UNIVERSITY OF ALBERTA

IDENTIFICATION AND CHARACTERIZATION OF TWO MAMMALIAN CENTROMERIC SATELLITE DNA FAMILIES

BY

CHARLES LEE ©

A Thesis

submitted to the Faculty of Graduate Studies and Research

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

IN

MEDICAL SCIENCES -

LABORATORY MEDICINE AND PATHOLOGY

EDMONTON, ALBERTA

FALL, 1996

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-612-18064-6

Canada

UNIVERSITY OF ALBERTA
RELEASE FORM

NAME OF AUTHOR: **CHARLES LEE**

TITLE OF THESIS: **IDENTIFICATION AND CHARACTERIZATION OF TWO MAMMALIAN CENTROMERIC SATELLITE DNA FAMILIES**

DEGREE **DOCTOR OF PHILOSOPHY**

YEAR THIS DEGREE GRANTED **1996**

Permission is hereby granted to the **UNIVERSITY OF ALBERTA LIBRARY** to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Charles Lee

Charles Lee
#210A, 3815-107 Street
Edmonton, Alberta, Canada
T6J 2N7

Date: October 3, 1996

University of Alberta
Edmonton

Canada T6G 2R7

Department of Laboratory Medicine and Pathology
Faculty of Medicine

4B1 WC Mackenzie Health Sciences Centre
Telephone (403) 492-8806

UNIVERSITY OF ALBERTA

RELEASE FORM

As a co-author of the article entitled *Human gamma X satellite DNA: an X chromosome specific centromeric DNA sequence*, published in *Chromosoma* 104: 103-112, 1995, I hereby give permission to Charles Lee, of 210A-3815-107 Street, Edmonton, Alberta, Canada, to use any version of this article as a portion of his Ph.D. thesis at the University of Alberta.

Xiang Li

Xiang Li, Ph.D.
Senior Scientist, Molecular Diagnostics
Smith Kline Beecham
709 Swedeland Road, P.O. Box 1539
King of Prussia, PA, U.S.A.
19406

Date: 4/17/96

University of Alberta
Edmonton
Canada T6G 2R7

Department of Laboratory Medicine and Pathology
Faculty of Medicine
4B1 WC Mackenzie Health Sciences Centre
Telephone (403) 492-8806

UNIVERSITY OF ALBERTA

RELEASE FORM

As a co-author of the article entitled *Human gamma X satellite DNA: an X chromosome specific centromeric DNA sequence*, published in *Chromosoma* 104: 103-112, 1995, I hereby give permission to Charles Lee, of 210A-3815-107 Street, Edmonton, Alberta, Canada, to use any version of this article as a portion of his Ph.D. thesis at the University of Alberta.

E.W. Jabs, M.D.
Associate Professor
Center for Medical Genetics, 10-04 CMSC
Johns Hopkins University School of Medicine
600 N Wolfe Street
Baltimore, MD, U.S.A. 21287-3914

Date: 4/22/96

University of Alberta
Edmonton

Canada T6G 2R7

Department of Laboratory Medicine and Pathology
Faculty of Medicine

4B1 WC Mackenzie Health Sciences Centre
Telephone (403) 492-8806

UNIVERSITY OF ALBERTA

RELEASE FORM

As a co-author of the article entitled *Human gamma X satellite DNA: an X chromosome specific centromeric DNA sequence*, published in *Chromosoma* 104: 103-112, 1995, I hereby give permission to Charles Lee, of 210A-3815-107 Street, Edmonton, Alberta, Canada, to use any version of this article as a portion of his Ph.D. thesis at the University of Alberta.

C.C. Lin

C.C. Lin, Ph.D.
Professor, Laboratory Medicine & Pathology
5B4.49 W.C. Mackenzie Health Sciences Centre
University of Alberta
Edmonton, Alberta, Canada
T6G 2B7

Date: April 2, 1996

University of Alberta
Edmonton

Department of Laboratory Medicine and Pathology
Faculty of Medicine

Canada T6G 2R7

4B1 WC Mackenzie Health Sciences Centre
Telephone (403) 492-8806

UNIVERSITY OF ALBERTA

RELEASE FORM

As a co-author of the article entitled *Human gamma X satellite DNA: an X chromosome specific centromeric DNA sequence*, published in *Chromosoma* 104: 103-112, 1995, I hereby give permission to Charles Lee, of 210A-3815-107 Street, Edmonton, Alberta, Canada, to use any version of this article as a portion of his Ph.D. thesis at the University of Alberta.

Dean Court
D.R. Court, B.Sc.
1855-37 Street
Edmonton, Alberta, Canada
T6L 3G8

Date: April 2, 1996

University of Alberta
Edmonton

Department of Laboratory Medicine and Pathology
Faculty of Medicine

Canada T6G 2R7

4B1 WC Mackenzie Health Sciences Centre
Telephone (403) 492-8806

UNIVERSITY OF ALBERTA

RELEASE FORM

As a co-author of the article entitled *Conservation of a 31 bp bovine subrepeat in centromeric satellite DNA monomers of Cervus elaphus and other cervid species*, currently in press in *Chromosome Research*. I hereby give permission to Charles Lee, of 210A-3815-107 Street, Edmonton, Alberta, Canada, to use any version of this article as a portion of his Ph.D. thesis at the University of Alberta.

CC Lin

C.C. Lin, Ph.D.
Professor, Laboratory Medicine & Pathology
5B4.49 W.C. Mackenzie Health Sciences Centre
University of Alberta
Edmonton, Alberta
T6G 2B7

Date: April 2, 1996

University of Alberta
Edmonton

Canada T6G 2R7

Department of Laboratory Medicine and Pathology
Faculty of Medicine

4B1 WC Mackenzie Health Sciences Centre
Telephone (403) 492-8806

UNIVERSITY OF ALBERTA

RELEASE FORM

As a co-author of the article entitled Higher-order organization of subrepeats and the evolution of cervid satellite I DNA.

submitted to J. Molecular Evolution, I hereby give permission to Charles Lee, of 210A-3815-107 Street, Edmonton, Alberta, Canada, to use any version of this article as a portion of his Ph.D. thesis at the University of Alberta.

Dean Court

D.R. Court, B.Sc.
1855-37 Street
Edmonton, Alberta, Canada
T6L 3G8

Date: April 3/1996

University of Alberta
Edmonton

Canada T6G 2R7

Department of Laboratory Medicine and Pathology
Faculty of Medicine

4B1 WC Mackenzie Health Sciences Centre
Telephone (403) 492-8806

UNIVERSITY OF ALBERTA

RELEASE FORM

As a co-author of the article entitled Higher-order organization of subrepeats and the evolution of cervid satellite I DNA submitted to J. Molecular Evolution, I hereby give permission to Charles Lee, of 210A-3815-107 Street, Edmonton, Alberta, Canada, to use any version of this article as a portion of his Ph.D. thesis at the University of Alberta.

C.C. Lin

C.C. Lin, Ph.D.
Professor, Laboratory Medicine & Pathology
5B4.49 W.C. Mackenzie Health Sciences Centre
University of Alberta
Edmonton, Alberta
T6G 2B7

Date: April 3, 1996

University of Alberta
Edmonton

Department of Laboratory Medicine and Pathology
Faculty of Medicine

Canada T6G 2R7

4B1 WC Mackenzie Health Sciences Centre
Telephone (403) 492-8806

UNIVERSITY OF ALBERTA

RELEASE FORM

As a co-author of the article entitled Higher-order organization of subrepeats and the evolution of cervid satellite I DNA submitted to J. Molecular Evolution, I hereby give permission to Charles Lee, of 210A-3815-107 Street, Edmonton, Alberta, Canada, to use any version of this article as a portion of his Ph.D. thesis at the University of Alberta.

Jennifer Haslett
#317, Range Lake Manor South
492 Range Lake Road
Yellowknife, NWT
X1A 3G4

Date: Apr. 3 1996

University of Alberta
Edmonton

Department of Laboratory Medicine and Pathology
Faculty of Medicine

Canada T6G 2R7

4B1 WC Mackenzie Health Sciences Centre
Telephone (403) 492-8806

UNIVERSITY OF ALBERTA

RELEASE FORM

As a co-author of the article entitled Higher-order organization of subrepeats and the evolution of cervid satellite I DNA submitted to J. Molecular Evolution, I hereby give permission to Charles Lee, of 210A-3815-107 Street, Edmonton, Alberta, Canada, to use any version of this article as a portion of his Ph.D. thesis at the University of Alberta.

Charles Cho
Charles Cho
1215-115A Street
Edmonton, Alberta, Canada
T6J 6Z3

Date: April 3, 1996

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommended to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **IDENTIFICATION AND CHARACTERIZATION OF TWO MAMMALIAN CENTROMERIC SATELLITE DNA FAMILIES** submitted by **CHARLES LEE** in partial fulfillment of the requirements for the degree of **DOCTOR OF PHILOSOPHY** in Medical Sciences - Laboratory Medicine and Pathology.

C.C. Lin

Dr. C.C. Lin
Supervisor
Department of Laboratory Medicine
and Pathology

Dr. D.B.C. Ritchie
Supervisory Committee Member
Department of Medicine

Dr. J.B. Rattner
Supervisory Committee Member
Department of Anatomy
and Medical Biochemistry
University of Calgary

Dr. R.A. Stinson
Examination Chair and Examiner
Department of Laboratory Medicine
and Pathology

J. Kuspira

Dr. J. Kuspira
Supervisory Committee Member
Department of Genetics

Dr. G.D. Burkholder
External Examiner
Department of Anatomy
and Cell Biology
University of Saskatchewan

Date: **September 26, 1996**

FOR MY MENTOR, PROFESSOR C.C. LIN.

Abstract

This thesis describes the characterization of two mammalian centromeric satellite DNA families, human gamma satellite DNA and cervid satellite I DNA. A subfamily of human gamma satellite DNA, specific for the centromere of the X chromosome, was identified and defined. This subfamily is characterized by tandemly organized, 220 bp, GC-rich repeat units (monomers) which share approximately 62% sequence similarity to the repeat units of the previously described gamma 8 satellite DNA subfamily. A conserved 20 bp continuous DNA region was observed in the monomers of gamma 8 and X satellite DNAs. The gamma X satellite DNA subfamily comprises approximately 0.5 Mb of the X chromosome centromere. Pulsed field gel electrophoresis studies suggest that restriction fragment lengths of this DNA subfamily are well conserved among unrelated individuals. The fluorescent *in situ* hybridization (FISH) localization pattern of this DNA subfamily are often seen as two distinct fluorescent dots at the lateral sides of the primary constriction. This hybridization pattern suggests a close proximity of gamma satellite DNA to the kinetochore region and implies a functional role for this DNA sequence in proper chromosome segregation. Further studies on the functional significance of this DNA sequence include the observed conservation of gamma satellite DNA in the genomes of two old world primates (African green monkey and chimanzee). On the other hand, gamma satellite DNA could not be detected in a mitotically stable marker chromosome using the currently available detection sensitivity of FISH.

Cervid satellite I DNA is a prominent centromeric satellite DNA family which is well conserved in deer species. Over 10,000 nucleotide bases of cervid centromeric satellite I DNA, from 10 independent clones of 6 different deer species, was sequenced and subjected to critical sequence analyses and comparisons. From this work, it was postulated that this DNA family originated from the amplification of a 31 bp DNA sequence (also found in the centromeric DNA of bovine species) which produced a 0.8 kb higher-order

repeat unit. The divergence of telemetacarpalial deer from plesiometacarpalial deer is associated with a 0.18 kb tandem duplication in this higher-order repeat unit, resulting in the genomic organization of this DNA family as 1 kb higher-order repeats in the former palentological division and 0.8 kb higher-order repeats in the latter palentological division. These higher order repeats have maintained more than 95% sequence similarity between centromeric repeats on nonhomologous chromosomes of a given animal, as well as between different animals of a particular deer species. Such data suggest the presence of functional constraints on these centromeric DNAs to maintain such sequence conservation and explain how different sized monomers can still belong to the same DNA family.

Acknowledgements

I am deeply indebted to my supervisor, Dr. C.C. Lin, for his consistent encouragement, enthusiastic guidance, and valued friendship throughout the last six years. As a result, I have thoroughly enjoyed my graduate studies and become the envy of many fellow graduate students. My appreciations also go out to my supervisory committee members (Drs. J.B. Rattner, J. Kuspira, and D.B.C. Ritchie) who have always made themselves available to me for valuable advice and assistance, despite their busy schedules. As well, my thanks to Drs. N. Telford, R.A. Stinson, and G.D. Burkholder for their time and expertise in making my candidacy and final oral defenses challenging, yet enjoyable.

This Ph.D. program could not have been completed without the help and friendships of numerous individuals. Much of the pulsed field gel work and analyses were made possible during my training with Dr. E.W. Jabs at the Johns Hopkins University School of Medicine. At Johns Hopkins, I also received support from Drs. Xiang Li and Amy Lewanda. In the Cytogenetics Research lab, University of Alberta, much thanks goes out to Dean Court and David McFaddyen for their friendships and help with experiments and write ups. Also, my sincere gratitude goes out to the staff of the Clinical Cytogenetics Laboratory (University of Alberta Hospitals), certain members of the Department of Genetics (University of Alberta), and Eastwood Management. Last, but definately not least, a special thanks to our wonderful neighbors: the labs of Drs. Mike Walter, Ian MacDonald, and Rachel Wevrick, for providing me with a home away from home. (I always looked forward to coming in to the lab knowing that you were all there!)

I gratefully acknowledge the financial assistance of a 75th Anniversary University of Alberta - Faculty of Medicine Graduate Student Award and an Alberta Heritage Foundation for Medical Research (AHFMR) Studentship. All research conducted during the course of my Ph.D. studies were also made possible by an operating grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada to Dr. C.C. Lin.

Finally, I thank my wife, Cecilia, my father, mother, James, and Jenny for their love and patience, especially during the difficult times, and my Lord, Jesus Christ, for blessing me so abundantly and for opening this door in my life.

Table of Contents **Page**

---

1. Introduction. 1

Repetitive DNA in mammalian genomes. 2

Human centromeric DNAs. 4

Satellite DNA families. 4

1. Human classical satellite DNAs. 4

2. Alpha satellite DNA. 9

3. Beta satellite DNA. 15

4. Gamma satellite DNA. 16

5. 48 bp satellite DNA. 17

6. Sn5 satellite DNA. 20

Interspersed repetitive DNA in human centromeres. 20

1. SINES. 20

2. LINES. 21

3. 724 Sequence Family. 21

Unclassified human centromeric repetitive DNA sequences. 22

1. HaeIII DNA Family. 22

2. ATRS DNA. 22

3. Centromeric repetitive elements from the Y chromosome. 23

Cervid Centromeric DNAs. 23

Repetitive DNA and the karyotypic evolution in the Asian muntjacs. 24

Major cervid centromeric satellite DNA (Cervid satellite I DNA). 26

Rationale and scope of this thesis. 27

Page

2. Discovery and preliminary characterization of human gamma X satellite DNA. 45

Introduction. 46

Materials and Methods. 48

Isolation of the 2D12/E2 subclone. 48

Southern blot hybridization. 48

DNA sequencing. 49

Copy number determination. 49

Chromosome preparations. 50

Fluorescence *in situ* hybridization. 50

Results. 51

Restriction enzyme analysis of the 2D12/E2 clone. 51

DNA sequence analysis of 2D12/E2. 56

Copy number estimation. 61

Fluorescence *in situ* hybridization. 61

Discussion. 66

Lack of macro restriction fragment length polymorphisms (RFLPs) with HpaI and SstI. 66

Inverted repeats and a conserved 20 bp region. 66

Estimation of copy number and array length. 67

Localization and organization of gamma satellite DNA. 68

Page

3. Ascertaining the presence of gamma satellite DNA in higher primates and a mitotically stable marker chromosome. 75

Prologue. 76

Introduction. 77

Materials and Methods. 80

Detection of gamma satellite DNA in two primate species. 80

Southern blot hybridization. 80

Fluorescence *in situ* hybridization. 81

Detection of gamma satellite DNA in a mitotically stable marker chromosome. 81

Results. 81

Presence of gamma satellite DNAs in two primate species. 81

Southern blot hybridization. 81

Fluorescence *in situ* hybridization. 84

Apparant absence of gamma satellite DNA in a mitotically stable marker chromosome. 84

Discussion. 87

Gamma satellite DNA are present in the genomes of the African green monkey and chimpanzee. 87

Gamma satellite DNA is undetectable in a mitotically stable human marker chromosome. 90

4. Conservation of a 31 bp bovine subrepeat in cervid centromeric satellite DNA. 97
Prologue. 98
Introduction. 99
Materials and Methods. 100
Cloning of a centromeric satellite DNA monomer from the European red deer. 100
Characterization of the Ce-Pst1 clone. 101
1. Southern blot analyses. 101
2. Fluorescence *in situ* hybridization. 101
3. DNA sequencing. 102
Source of other cervid centromeric satellite DNA monomer sequences. 102
Internal repeat analyses. 102
Results. 104
Characterization of the European red deer centromeric satellite DNA clone, Ce-Pst1. 104
An internal 31 bp periodicity for each deer monomer clone. 111
Discussion. 125
Ce-Pst1 and a previously isolated 770 bp red deer clone. 125
CAGG and GAGG tetramers in the Ce-Pst1 DNA sequence. 126
31 bp subrepeats in the cervid monomers. 127
Evolution of cervid centromeric satellite DNA. 127

Page

---

5. Organization and evolution of cervid satellite I DNA. 134

Prologue. 135

Introduction. 136

Materials and Methods. 137

Source of genomic DNAs. 137

Cloning and characterization of cervid satellite DNAs. 137

DNA sequence analyses and comparisons. 138

Results. 139

Initial characterization of newly isolated cervid satellite I DNA clones. 139

DNA sequence analyses of cervid satellite I DNA. 145

Intragenomic sequence conservation. 151

Intraspecific sequence conservation. 151

Interspecific sequence conservation. 157

Discussion. 157

Presence and organization of 31 bp subrepeats in cervid satellite I monomer clones. 157

Genesis of 0.8 kb and 1 kb cervid satellite I DNA monomers. 161

Implications of intraspecific sequence conservation. 162

6. Conclusions. 170

List of Tables **Page**

1.1 Human centromeric DNAs. 5

1.2 Evolution of alphoid monomer types. 11

1.3 Definition of four alphoid suprachromosomal families. 12

4.1 Size increases of in-frame peaks. 117

5.1 Satellite DNA clones isolated from 5 deer species. 140

5.2 Percent sequence divergence of 31 bp subrepeats in 1 kb monomers from mule deer. 150

5.3 Intragenomic and intraspecific sequence similarities. 156

5.4 Interspecific sequence homology among 10 representative centromeric satellite DNA clones. 160

List of Figures **Page**

1.1 Genomic map of the relative size and distribution of identified repetitive DNA sequences in the centromeres of human chromosomes 10 and Y. 19

2.1 Southern blot hybridization of human genomics DNAs, fractionated by conventional gel electrophoresis, and probed with clone 2D12/E2. 53

2.2 Southern blot hybridization of human genomic DNAs from 8 unrelated individuals, fractionated by pulsed field gel electrophoresis, and probed with clone 2D12/E2 or 50E4. 55

2.3 Nucleotide sequence and monomer organization of the 2D12/E2 insert. 58

2.4 Derivation of a 220 bp consensus sequence from the monomers in clone 2D12/E2. 60

2.5 Comparison of the 2D12/E2 monomer consensus sequence with the gamma 8 monomer consensus sequence. 63

2.6 Localization of the 2D12/E2 DNA probe to human male metaphase chromosomes. 65

3.1 Three functional domains of human centromeres. 79

3.2 Southern blot hybridization of gamma X satellite DNA to genomic DNAs from two old world monkeys. 83

3.3 Localization of human gamma 8 and X satellite DNAs to metaphase chromosomes of a male African green monkey and a male chimpanzee. 86

3.4 Localization of gamma 8 satellite DNA to metaphase chromosomes from a lymphoblast cell line containing a single mitotically stable chromosome 8 derived marker chromosome. 89

3.5 A proposed mechanism for the formation of the marker chromosome from patient AS. 92

4.1 Tandemly repetitive organization of the Ce-Pst1 DNA sequence in the genome of the European red deer (*Cervus elaphus hippelaphus*). 106

4.2 Localization of the Ce-Pst1 satellite DNA to the chromosomes of a male European red deer by fluorescence *in situ* hybridization. 108

4.3 DNA sequence and comparison strategy for the Ce-Pst clone. 110

4.4 The existance of a 31 bp periodicity in cervid centromeric satellite DNA clones: Ce-Pst1, C5, CCSatI, Rt-Pst3. 113

4.5 Derivation of consensus sequences for the 31 bp subrepeats in clones Ce-Pst1,C5, CCSatI, and Rt-Pst3. 119

Page

4.6 DNA sequence analyses of the monomer-specific 31 bp consensus sequences. 124

5.1 Nucleotide sequences of 9 monomer clones from 5 deer species. 142

5.2 Consensus sequences from 31 bp subrepeats of representative cervid monomer clones. 147

5.3 A 0.18 kb tandem duplication in 1 kb monomers of Oh-Msp1 and Oh-Msp2a, having a higher-order organization of 31 bp subrepeats. 153

5.4 Intragenomic and intraspecific monomeric sequence comparisons. 155

5.5 Alignment strategies for interspecific sequence comparisons of 10 representative cervid centromeric satellite I DNA monomer clones. 159

5.6 Proposed genesis of satellite I DNA in cervids. 164

5.7 Taxonomic reclassification of 10 deer species from which centromeric satellite I DNA clones have been isolated. 167

List of Abbreviations

0.1x SSC	- 15 mM NaCl, 1.5 mM sodium citrate; pH = 7.0
2x SSC	- 300 mM NaCl, 30 mM sodium citrate; pH = 7.0
BN Buffer	- 0.1M $NaHCO_3$, 0.05% (v/v) NP-40; pH = 8.0
bp	- base pair
CENP	- centromeric protein
DMEM	- Dulbecco minimal essential medium
EBV	- Epstein barr virus
EDTA	- ethylenediaminetetraacetic acid
FISH	- fluorescence *in situ* hybridization
h	- hour(s)
HMG	- high mobility group
HOR	- higher order repeat
kb	- kilobase
Mb	- megabase
min	- minute(s)
PBS	- phosphate buffered saline, pH = 7.5
RFLP	- restriction fragment length polymorphism
SDS	- sodium dodecyl sulphate
s	- seconds
SV40	- simian virus 40
TE	-Tris-HCl / EDTA
y	- year

'"Given a sufficient lack of comprehension, anything (and that includes a quartet of Mozart) can be declared to be junk. (Repetitive DNA) is, in fact, a collectors' item.'"

Zuckerlandl E, Henning W (1995) Tracking heterochromatin. *Chromosoma* 104: 75.

CHAPTER 1

INTRODUCTION

A literature review of centromeric DNAs in human and cervid species.

The term centromere is thought to have first been coined by Waldeyer in 1903. It is the chromosome structure responsible for the proper segregation of chromosomes during mitosis and meiosis. Acentric chromosomes (chromosomes which lack a functional centormere) experience chromosome lag during cell division and are subsequently lost in the resulting daughter cells. Although centromeres can be non localized in the chromosomes of certain eukaryotes (e.g. Hughes-Schrader and Ris 1941, Bokhari and Godard 1980, Albertson and Thomson 1982), most centromeres are localized in nature, appearing as a primary constriction at a fixed locus for each nonhomologous chromosome. For the purposes of this introductory chapter, centromeres will refer to the localized type.

The centromeric locus of mammalian chromosomes are thought to be composed of complexes of centromeric proteins and repetitive DNAs. Upon this specific organization of protein/DNA complexes, a proteinaceous structure (known as the kinetochore) is formed. The specific DNA sequences required for a functioning centromere in mammalian cells has not yet been ascertained and is the focus of this thesis.

Repetitive DNA in mammalian genomes.

Higher eukaryote genomes contain various repetitive DNA sequences which do not appear to code for proteins. The discovery of repetitive sequences originated from an unexpected observation made when high molecular weight DNA was purified in cesium chloride centrifugation gradients. In addition to the expected main band which contained the bulk of genomic DNA, Kit (1961) observed additional small discrete (satellite) bands. DNA reassociation experiments revealed that these "satellite" bands had DNA which reannealed much more rapidly than the DNA of the main band (Waring and Britten 1966). From this and other studies, Britten and Kohne (1968) concluded that satellite DNA bands consisted of repeated DNAs, a ubiquitous component of higher eukaryote genomes.

These non-coding repetitive DNAs generally exhibit two forms of organization. Repeat units can either be interspersed among other DNA sequences or tandemly arranged in continuous DNA arrays. The latter scenario is characteristic of repetitive DNAs which have been termed "satellite DNAs"[1].

For some time, satellite DNAs were considered non-essential to higher eukaryote genomes and even coined by some as "junk" DNA (Doolittle and Sapienza 1980, Orgel and Crick 1980). However, the frequent localization of these "junk" DNAs to functional chromosome domains, such as centromeres, remained an enigma. Some arguments were made that the extremely low rate of recombination at constitutive centromeric heterochromatin[2] regions could account for the persistence of repetitive DNA sequences in these chromosome regions (Charlesworth *et al.* 1986). Others speculated that satellite DNAs may exist to provide certain cellular functions without necessarily coding for proteins. Such cellular functions could include the provision of binding sites for specific nuclear proteins or the adoption of irregular, locus specific secondary structures for transcriptional regulation of adjacent protein coding sequences (reviewed in Vogt 1990). Some centromeric satellite DNAs were also believed to assist in recognition of homologues during meiotic pairing (Haaf *et al.* 1986, Choo *et al.* 1988); an idea recently shown to be valid in *Drosophila* (Karpen *et al.* 1996).

To investigate the nature and function of centromeric satellite DNAs, which may be responsible for centromere activity in higher eukaryotes, numerous studies have been directed at identifying and characterizing DNA sequences in centromeres. Each centromeric satellite DNA family is composed of repeat units (monomers) which are often identified by a characteristic ladder pattern of restriction fragments produced by cleavage at a regular periodicity. The sequence and size of independently isolated monomers are similar, but seldom identical and can sometimes be organized in a hierarchical fashion to produce chromosome-specific higher-order repeats (HORs), comprised of integral number of monomers. These HORs are thought to reflect recent multimeric units of amplification and exhibit increased sequence similarity, on the order of > 95% between related HORs (e.g. see Willard and Waye 1987b for a review).

---

[1] Pech *et al.* (1979) suggested that any tandem repetition of a unit DNA sequence be considered as "satellite DNA".

[2] Constitutive heterochromatin refers to chromosome regions which are thought to be densely packaged and often associated with repetitive DNA sequences.

Among higher eukaryote species, the nucleotide composition of human centromeres has been the most extensively characterized. Comparatively less is known of the centromeric satellite DNA of deer. However, some interesting findings on the major cervid centromeric satellite DNA family have recently emerged. Following is a comprehensive review of centromeric DNAs currently identified in these two mammalian species.

Human centromeric DNAs.

All currently identified human centromeric DNAs are repetitive in nature, with the vast majority of these DNA sequences being tandemly repeated. Extensive reference will not be made to human centromeric repetitive DNAs which have not been characterized at the level of the primary nucleotide sequence. A list of human centromeric DNAs, which have been sequenced and localized to the centromeres of particular chromosomes, is provided in Table 1.1.

Satellite DNA families.

1. Human classical satellite DNAs. The first human centromeric DNAs to be identified were isolated from three genomic DNA fractions having slightly different buoyant densities in $CsSO_4$ gradients supplemented with DNA-binding metals such as silver or mercury (Corneo *et al.* 1967, 1968, 1970, 1971). These DNA fractions were referred to as satellites I, II, and III. Each fraction most likely contains a heterogeneous population of repeated DNA sequences with similar buoyant densities, complicating the analysis of these satellite DNA fractions. As it would not be appropriate to define these DNA fractions as individual satellite DNA families, Prosser *et al.* (1986) identified a single predominant family of simple repeated sequence in each classical satellite DNA fraction. These three simple sequence families were named satellite DNA families 1, 2, and 3, to indicate their

TABLE 1.1. Human centromeric DNAs

Chromosome Region	Repeat Unit (bp)	Repetitive DNA family	Clone(s)	HOR (kb/enzyme)	Array size estimation (Mb)	References
1 cen	171	Alpha Sat.	pSD1-1	1.9 / HindIII	0.44 - 1.51	Waye *et al.* 1987c, Wevrick and Willard 1989
	n.d.	ATRS	p16-1	n.d.	n.d.	Wevrick *et al.* 1992
2 cen	5, 3	Sat. 2	S.2	n.d.	n.d.	Tagarro *et al.* 1994a
	171	Alpha Sat.	pBS4D, p2-11	0.68 / XbaI	1.05 - 2.9	Rocchi *et al.* 1990, Haaf and Willard 1992
	> 406	sn5	pcrsn5	n.d.	n.d.	Johnson *et al.* 1992
	n.d.	ATRS	p16-1	n.d.	n.d.	Wevrick *et al.* 1992
3 cen	17, 25	Sat. 1	pTRI-6	3 / SspI	0.2 - 2	Kalitsis *et al.* 1993
	171	Alpha Sat.	VII B4, p3-9	2.75 / HindIII	1 - >1.5	Delattre *et al.* 1988, Waye and Willard 1989b
4 cen	17, 25	Sat. 1	pTRI-6	3 / SspI	0.2 - 2	Kalitsis *et al.* 1993
	171	Alpha Sat.	p4n1/4	3.4 / SacI	n.d.	D'Aiuto *et al.* 1993
	171	Alpha Sat.	pYAM3-84	2.6, 0.6 / MspI	n.d.	Mashkova *et al.* 1994
5 cen	171	Alpha Sat.	pG-A16*	2.25 / EcoRI	0.17 - 0.34	Hulsebos *et al.* 1988
6 cen	171	Alpha Sat.	p308	3 / BamHI	~ 2.3	Jabs *et al.* 1984, 1989
7 cen	171	Alpha Sat.	pα7d1, pα7t1	1 / EcoRI-partial	1.53 - 3.81	Waye *et al.* 1987b, Wevrick and Willard 1989
	171	Alpha Sat.	pMGB7	2.7 / HindIII	0.1 - 0.55	Waye *et al.* 1987b, Wevrick and Willard 1989
	n.d.	ATRS	p16-1	n.d.	n.d.	Wevrick *et al.* 1992
8 cen	171	Alpha Sat.	pJM128, pBS8-164	2.5 / HindIII	1.5 - 2.19	Donlon *et al.* 1987, Ge *et al.* 1992
	220	Gamma Sat.	50E1, 50E4	n.d.	~0.43	Lin *et al.* 1993
9 cen	68	Sau3A/Beta Sat.	pB3	2.5 / EcoRI	0.125 -0 .25	Waye and Willard 1989a, Greig and Willard 1992
	171	Alpha Sat.	pMR9A	2.7 / PstI	n.d.	Rocchi *et al.* 1991
	n.d.	ATRS	p16-1	n.d.	n.d.	Wevrick *et al.* 1992
10 cen	5, 3	Sat. 2	p375M2.4	1.8 / EcoRI	0.9	Jackson *et al.* 1993
	5, 10	Sat. 3	mC219.2/28	n.d.	0.52 - 0.62	Jackson *et al.* 1993
	171	Alpha Sat.	pα10RP8	1.35 / RsaI	1.39 - 2.515	Devilee *et al.* 1988, Wevrick and Willard 1989
11 cen	171	Alpha Sat.	pLC11A	0.85 / XbaI	1.96 - 4.76	Waye *et al.* 1987a, Wevrick and Willard 1989
12 cen	171	Alpha Sat.	pBR12, pSP12-1	1.4 / PvuII	2.25 - 4.3	Baldini *et al.* 1990, Greig *et al.* 1991
13 cen	5, 10	Sat. 3	pTR9-H2	2.6 / FokI	n.d.	Vissel *et al.* 1992
	17, 25	Sat. 1	pTRI-6	3 / SspI	0.2 - 2	Kalitsis *et al.* 1993
	171	Alpha Sat.	L1.26	n.d	~1	Devilee *et al.* 1986
	n.d.	ATRS	p16-1	n.d.	n.d.	Wevrick *et al.* 1992
	> 406	sn5	pcrsn5	n.d.	n.d.	Johnson *et al.* 1992

Legend:
cen - centromere
Sat. - Satellite
HOR - Higher order repeat
n.d. - not determined
* - Not shown to be chromosome specific

TABLE 1.1. Human centromeric DNAs (continued)

Chromosome Region	Repeat Unit (bp)	Repetitive DNA family	Clone(s)	HOR (kb/enzyme)	Array size estimation (Mb)	References
14 cen	5, 10	Sat. 3	pTR9-H2	2.6 / FokI	n.d.	Vissel *et al.* 1992
	17, 25	Sat. 1	pTRI-6	3 / SspI	0.2 - 2	Kalitsis *et al.* 1993
	171	Alpha Sat.	p82H	2.4 / EcoRI+HindIII	> 0.035	Waye *et al.* 1988
	> 406	sn5	pcrsn5	n.d.	n.d.	Johnson *et al.* 1992
	n.d.	ATRS	p16-1	n.d.	n.d.	Wevrick *et al.* 1992
15 cen	5, 10	Sat. 3	15C-3-1, pTR9-H2	n.d.	n.d.	Jabs *et al.* 1989, Vissel *et al.* 1992
	17, 25	Sat. 1	pTRI-6	3 / SspI	0.2 - 2	Kalitsis *et al.* 1993
	171	Alpha Sat.	pTRA-20	2.5 / DraI	n.d.	Choo *et al.* 1990b
	171	Alpha Sat.	pTRA-25	4.5 / EcoRI	n.d.	Choo *et al.* 1990b
	> 406	sn5	pcrsn5	n.d.	n.d.	Johnson *et al.* 1992
	n.d.	ATRS	p16-1	n.d.	n.d.	Wevrick *et al.* 1992
16 cen	171	Alpha Sat.	pSE16	1.7 / Sau3A	0.43 - 2	Greig *et al.* 1989, Wevrick and Willard 1989
	n.d.	ATRS	p16-1	n.d.	n.d.	Wevrick *et al.* 1992
17 cen	171	Alpha Sat.	p17H8, p3-6	2.7 / EcoRI	1.35 - 2.7	Waye and Willard 1986
	171	Alpha Sat.	E7	2.2 / PstI	~4.8	Choo *et al.* 1987
	n.d.	ATRS	p16-1	n.d.	n.d.	Wevrick *et al.* 1992
18 cen	171	Alpha Sat.	L1.84, pYAM 9-60	1.36 / EcoRI	~1.368	Devilee *et al.* 1986, Alexandrov *et al.* 1991
	171	Alpha Sat.	pYAM 4-22	1.7 / HindIII	n.d.	Alexandrov *et al.* 1991
19 cen	171	Alpha Sat.	pGA16*	2.25 / EcoRI	0.17 - 0.34	Hulsebos *et al.* 1988
20 cen	171	Alpha Sat.	pZ20	1 / HinfI	n.d.	Baldini *et al.* 1992
	> 406	sn5	pcrsn5	n.d.	n.d.	Johnson *et al.* 1992
21 cen	5, 10	Sat. 3	pTR9-H2	2.6 / FokI	n.d.	Vissel *et al.* 1992
	17, 25	Sat. 1	pTRI-6	3 / SspI	0.2 - 2	Kalitsis *et al.* 1993
	48	48 bp Sat.	pcos6	n.d.	0.2 -1.1	Mullenbach *et al.* 1992
	171	Alpha Sat.	L1.26	n.d	~1	Devilee *et al.* 1986
	> 406	sn5	pcrsn5	n.d.	n.d.	Johnson *et al.* 1992
	n.d.	ATRS	p16-1	n.d.	n.d.	Wevrick *et al.* 1992
22 cen	17, 25	Sat. 1	pTRI-6	3 / SspI	0.2 - 2	Kalitsis *et al.* 1993
	48	48 bp Sat.	p22hom48.4	n.d.	n.d.	Metzdorf *et al.* 1988
	171	Alpha Sat.	p22/1:0.73	2.1, 2.8 / EcoRI	n.d.	McDermid *et al.* 1986
	> 406	sn5	pcrsn5	n.d.	n.d.	Johnson *et al.* 1992
	n.d.	ATRS	p16-1	n.d.	n.d.	Wevrick *et al.* 1992
X cen	171	Alpha Sat.	pXBR-1	2 / BamHI	1.38 - 3.73	Yang *et al.* 1982, Mahtani and Willard 1990
Y cen	5,10	Sat. 3	pY5	n.d.	0.4	Cooper *et al.* 1993
	48	48 bp Sat.	pY48	n.d.	≤ 0.019	Cooper *et al.* 1992, 1993
	48	48 bp Sat.	pY48	n.d.	≤ 0.021	Cooper *et al.* 1992, 1993
	68	Sau3A/Beta Sat.	pKFC68	16.9 / HaeIII	~0.027	Cooper *et al.* 1992
	171	Alpha Sat.	cos Y84	5.5 / EcoRI	0.24 - 1.6	Wolfe *et al.* 1985, Larin *et al.* 1994
	171	Alpha Sat.	YII3.1	n.d.	~0.018	Cooper *et al.* 1993

Legend:
cen - centromere
Sat. - Satellite
HOR - Higher order repeat
n.d. - not determined
* - Not shown to be chromosome specific

enrichment in classical satellite fractions I, II, and III, respectively. Satellite DNA family 1 is comprised of 42 bp repeats arranged as two alternating DNA sequences (i.e. A-B-A-B-...). Sequence A is a 17 bp repeated sequence (ACATAAAATAT$\genfrac{}{}{0pt}{}{C}{G}$AAAGT) and sequence B is a 25 bp repeat unit (ACCCAAAAT$\genfrac{}{}{0pt}{}{A}{G}T\genfrac{}{}{0pt}{}{A}{G}$TATTATATACTGT). Satellite DNA family 2 consists of poorly conserved ATTCC repeats which contain the sequence (ATTCC ATTCG)$_2$, followed by one or two ATG(s) (Deininger *et al* 1981[3], Shaul *et al.* 1986, Jeanpierre 1994). Satellite DNA family 3 is defined by repetitions of well conserved ATTCC pentameric repeats, interspersed with a specific 10 bp sequence (A$\genfrac{}{}{0pt}{}{T}{C}$TCGGGTTG).

Initial *in situ* hybridization studies using the heterogeneous populations of repeated DNAs from each satellite DNA fraction suggested that the human classical satellite DNA fractions had both pericentromeric and non-centromeric chromosome localizations (e.g. Jones *et al.* 1973, Jones *et al.* 1974, Gosden *et al.* 1975, Prosser *et al.* 1981). However, interpretation of these *in situ* hybridization results should reflect the sequence heterogeneity of each DNA fraction. *In situ* hybridization experiments with clones of known sequences suggested that satellite 1 DNA is primarily localized to the pericentric regions of chromosomes 3, 4, as well as the pericentric regions and the short arms of the acrocentric chromosomes (Kalitsis *et al.* 1993, Meyne *et al.* 1994[4], Tagarro *et al.* 1994b[5]). Satellite 2 DNA sequences were shown to be primarily localized to the variable heterochromatic regions of chromosomes 1 (Cooke and Hindley 1979[6], Tagarro *et al.* 1994a), and 16 (Moyzis *et al.* 1987, Schwarzacher-Robinson *et al.* 1988), with less prominent domains in the pericentromeric regions of chromosomes 2 and 10 (Tagarro *et al.* 1994a). Satellite 3 DNA sequences were predominantly localized by *in situ* hybridization to the variable heterochromatin of chromosome 1 (Nakahori *et al.* 1986) and 9 (Moyzis *et al.* 1987, Schwarzacher-Robinson *et al.* 1988), and the long arm of the Y-chromosome (Cooke *et al.* 1982). Satellite 3 DNA sequences were also localized to the short arms of all acrocentric chromosomes, proximal to the rDNA region (Wang *et al.* 1984, Higgins *et al.* 1985, Choo *et al.* 1990a). Other *in situ* hybridization experiments suggested that arrays of satellite 3 may also exist in chromosomes 5, 10, 17, and 20 (Tagarro *et al.* 1994a). Grady *et al.* (1992) suggested that diverged pentameric repeats, similar to satellite 3, may actually exist at the centromeric regions of all human chromosomes. Molecular analyses have subsequently identified satellite 3 DNA sequences have been identified adjacent to alpha satellite DNA, in the centromeric regions of chromosomes 10 (Jackson *et al.* 1992, 1993), 13, 14, 15[7], and 21 (Jabs *et al.* 1989, Vissel *et al.* 1992).

---

[3] In this study clone pPD17 is referred to as satellite III, but based on the above definitions, is actually a human classical satellite 2 clone.

[4] This paper also demonstrated by strand specific hybridization methods (i.e. Chromosome Orientation -FISH) that satellite 1 sequences are arranged predominantly, if not completely, in a head-to-tail tandem fashion.

[5] Tagarro *et al.* (1994b) specifically suggested a chromosomal localization of 3q11.2 and 4q11-12 for satellite 1 sequences on chromosomes 3 and 4, respectively. They further revealed two satellite 1 domains in chromosome 13: one at 13p13 and a second at 13p11.2.

[6] In this communication, a 1.77 kb repeat (designated as satellite III) is shown by Southern blot hybridizations to DNA to originate from chromosome 1. However, a close examination of the partial sequence from clone λHS5, suggests that this 1.77 kb fragment is actually a satellite 2 clone.

It should also be mentioned that centromeric repetitive DNAs from other human satellite DNA fractions have also been reported. Corneo *et al.* (1972) identified a satellite DNA fraction IV, but based on similar buoyant densities, reassociation kinetics (Mitchell *et al.* 1979), and restriction digestion patterns (Frommer *et al.* 1982[8]), the repetitive sequences comprising human satellite DNA fraction IV are thought to be essentially identical to the repetitive DNAs of satellite DNA fraction III. A satellite DNA fraction, referred to as human satellite fraction C (Saunders *et al.* 1972, Chuang and Saunders 1974) had a buoyant density of 1.703 g/cm^3, unlike the DNA fractions identified by Corneo and coworkers. Repetitive sequences from this DNA fraction hybridized to the centromeric heterochromatin of the human acrocentric chromosomes as well as chromosome 9. Unfortunately, sequence data for the repetitive DNAs of this DNA fraction are not available.

---

[7] In the centromere of chromosome 15, identified satellite 3 sequences are comprised primarily of the diverged satellite 3 repeats: GGAAT or GGAGT.

[8] Disregarding the presence of alpha satellite DNA in fraction IV.

2. Alpha satellite DNA. Maio (1971) described the isolation of a highly repetitive DNA fraction in the African green monkey genome and referred to as component α. Gruss and Sauter (1975) demonstrated that this component α is composed mainly of 170 bp repeat units. Two years later, Rosenberg *et al.* (1977) sequenced a member of this α DNA family, which had somehow become incorporated into the genomes of certain SV40 viruses, rendering them defective. A simian alpha satellite DNA consensus sequence was then published from uncloned HindIII fragments of African Green monkey genomic DNA (Rosenberg *et al.* 1978). Manuelidis and Wu (1978) subsequently reported the consensus sequence of homologous repeats in uncloned human EcoRI DNA fragments, which shared 70% - 80% sequence identity with the African green monkey consensus sequence. More recently, consensus sequences for the 171 bp basic repeat of human alpha satellite DNA have been constructed with 130 independent monomers from 14 different human chromosomes (Willard and Waye 1987a) and with 293 independent monomers from the 24 different, non homologous human chromosomes (Choo *et al.* 1991).

Although individual alpha satellite DNA clones were used to detect alphoid DNA in the centromeric region of every human chromosome (Jabs *et al.* 1984[9], Mitchell *et al.* 1985), the 171 bp monomers display substantial intermonomer sequence divergence, on the order of 20% - 40%. Therefore, juxtaposed alphoid DNA monomers in one human centromere generally share no more sequence similarity than monomers from different chromosomes. However, most alphoid monomers are organized in a hierarchical fashion to produce chromosome-specific higher order repeats (HORs)[10]. Individual HORs which define each subfamily / subset have less than 5% sequence divergence (For a review, see Willard and Waye 1987b).

---

[9] Jabs *et al.* (1984) used clone 308 to show the presence of homologous sequences in the centromere of all human chromosomes (See note in added proof). However, clone 308 was not shown to contain alpha satellite DNA until it was actually sequenced (Jabs and Persico 1987).

[10] Alpha satellite DNA monomers in the human genome are not all organized in distinguishable hierarchical structures (e.g. Waye *et al.* 1988, Wevrick *et al.* 1992). In fact, Smith (1976) theorized that crossing over mechanisms would lead to less homogenized repeat units at the periphery of satellite DNA arrays, as compared to repeats towards the center of the array.

As for the genesis and evolution of alpha satellite DNA, the lack of any discernible subrepeat within alphoid monomers precludes the notion that this satellite DNA family was produced from amplification of a smaller DNA sequence. Based on studies of interchromosomal sequence similarities, Alexandrov *et al.* (1988, 1991, 1993) suggested that all human alpha satellite DNA could have arisen from two or three fundamentally different ancestral monomers, which in turn underwent amplification and sequence divergence to form 9 - 10 different monomer types (Table 1.2). These monomer types were then organized to produce four distinct alphoid "suprachromosomal families" (See Table 1.3). Suprachromosomal family 1 is characterized by a dimeric repeat unit, similar to that described by Wu and Manuelidis (1980), that leads to the formation of DNA arrays on chromosomes 1, 3, 5, 6, 7, 10, 12, 16, and 19. Suprachromosomal family 2 is characterized by another dimer, distinct from the dimer of suprachromosomal family 1, and leads to the genesis of DNA arrays on chromosomes 2, 4, 8, 9, 13, 14, 15, 18, 20, and 22. Suprachromosomal family 3 has a pentameric organization, distinctive of certain centromeric DNA arrays in chromosomes 1, 11, 17, and X. Suprachromosomal family 4 is characterized by a monomeric construction, similar to that found in African green monkey alpha satellite DNA which lacks identifiable HOR organization and exhibits increased intermonomeric sequence similarity. These monomers were found in the centromeric regions of the acrocentric chromosomes as well as the Y-chromosome.

From the data presented above, centromeres of certain human chromosomes contain more than one suprachromosomal family of alpha satellite DNA. Alexandrov *et al.* (1988) proposed that interchromosomal transfers could have contributed to the existence of more than one particular alphoid domain at a given centromere. Each alphoid domain could then have been subjected to very powerful homogenization processes mantaining more than 99% sequence identity between related HORs within DNA segments of 10 kb - 40 kb (Durfy and Willard 1989). Such homogenization processes may result from multiple recombination events, such as unequal crossing overs and sequence conversions.

Table 1.2. Evolution of alphoid monomer types

Ancestral Monomer	Monomer types	Reference
1	J1, D1, W1, W3	Alexandrov et al. 1991
2	J2, D2, W4, W5	
3	W2	
1	J2, D1, W1, W2, W3	Alexandrov et al. 1993
2	J1, D2, W4, W5, M1	

Table 1.3. Definition of four alphoid suprachromosomal families

Suprachromosomal Family	Chromosomes	Monomer Type	Repeat Organization
1	1,3,5,6,7,10,12, 16,19	J Type	...-J1-J2-... (Dimer)
2	2,4,8,9,13,14, 15,18,20,22	D Type	...-D1-D2-... (Dimer)
3	1,11,17, X	W Type	...-W1-W2- W3-W4-W5-... (Pentamer)
4	13, 14, 15, 21, 22, Y	M Type	...-M1-... (Monomer)

Note: Chromosomes having alphoid domains from more than one suprachromosomal family are underlined.

Alpha satellite DNA is primarily organized in a head-to-tail fashion in each DNA array. Albeit rare, a small number of inversions have been reported in alpha satellite DNA clones (Jabs and Persico 1987, Cooper *et al.* 1993, Bayne *et al.* 1994) and from PCR amplification analyses of genomic DNAs (Wevrick *et al.* 1992). Furthermore, alpha satellite DNA is not uniformly distributed between the centromeres of different human chromosomes, varying from approximately 250 kb[11] in the Y-chromosome (Cooper *et al.* 1993, Larin *et al.* 1994) to almost 5000 kb in chromosome 11 (Wevrick and Willard 1989). Polymorphic variation can also be observed in alpha satellite DNA arrays of homologous chromosomes. Enzymatic digestions, which cut rarely within alpha satellite DNA arrays, usually yield restriction fragment length polymorphisms (RFLPs) that display a Mendelian inheritance pattern (e.g. Jabs *et al.* 1989, Wevrick and Willard 1989, Haaf and Willard 1992) allowing the exploitation of alphoid DNAs as genetic markers.

Protein binding studies report a number of nuclear proteins which have an affinity for alpha satellite DNA. The most extensively characterized is CENP-B (CENtromeric Protein B) which was first identified as an 80 kDa polypeptide (Earnshaw and Rothfield 1985) but later thought to be closer to 65 kDa (Earnshaw *et al.* 1987). This protein has four main domains including two highly acidic regions, a helix-loop-helix DNA binding motif, and a proline-rich "hinge" region. CENP-B may form a complex comprising two alpha satellite molecules and dimerized CENP-B (Muro *et al.* 1992, Yoda *et al.* 1993). Masumoto *et al.* (1989) found that the CENP-B protein binds to a specific 17 bp motif (5'-CTTCGTTGGAAACGGGA-3'), now commonly referred to as the CENP-B box, in both human alpha satellite and mouse minor satellite DNA (Wong and Rattner 1988). Earnshaw *et al.* (1987) estimated 20,000 - 50,000 CENP-B molecules in every cell which bound to alphoid DNA in a stoichiometric fashion. Also consistent with the binding of CENP-B to certain alpha satellite DNA monomers is the distribution of CENP-B protein throughout the centromeric heterochromatin, beneath the kinetochore (Cooke *et al.* 1990).

---

[11] This estimate represents the combined length of two distinct alphoid arrays in the human Y chromosome.

The presence of CENP-B is not necessarily indicative of centromere activity. Immunofluorescent studies failed to detect CENP-B in the human Y-chromosome (Earnshaw *et al.* 1989, Pluta *et al.* 1990) and alphoid clones isolated from human Y-chromosomes all lack CENP-B boxes (Tyler-Smith and Brown 1987). Furthermore, CENP-B was detected at both functional and non-functional centromeres of stable dicentric chromosomes (Earnshaw *et al.* 1989). Nevertheless, CENP-B is thought to be a structural protein possibly contributing to a nucleosome phasing in centromeric heterochromatin (Widom and Klug 1985).

Strauss and Varshavsky (1984) isolated an abundant High Mobility Group (HMG)-like protein (called α-protein) from African green monkey cells which exhibits preferential binding to three specific sites in a single alpha satellite DNA monomer. This 10 kDa protein (Reeves and Nissen 1990[12]) recognizes the minor groove of a B-DNA double helix where runs of six or more A-T bp are present (Solomon *et al.* 1986). Solomon *et al.* (1986) noted that α-protein is identical to HMG-I protein (Lund *et al.* 1985) and is therefore now considered a member of the HMG-I family. Mutation experiments in Drosophila HMG-I like centromeric heterochromatin proteins suggested that these proteins may in fact contribute to higher levels of heterochromatin condensation and subsequent kinetochore nucleation (reviewed in Schulman and Bloom 1991).

Gaff *et al.* (1994) described a third putative alpha satellite binding protein, pJα, which has a molecular weight of 10 -15 kDa and has a specific affinity for a 9 bp alphoid-derived direct repeat (GTGAAAAAG), found at the junction of alpha satellite DNA and classical satellite 3 in chromosomes 13, 14, and 21. The study could not formally exclude the possibility that pJα is actually CENP-A, a 17 kDa centromere-specific histone (Earnshaw and Rothfield 1985) distantly related to histone H3 (Palmer *et al.* 1991, Sullivan *et al.* 1994). However, pJα is not thought to be an HMG-I protein based on competitive binding results with poly [d(I-C)] molecules and an oligonucleotide harbouring an HMG-I binding site (Gaff *et al.* 1994).

---

[12] Reeves and Nissen (1990) predicted a molecular structure for the three conserved binding regions of α protein and found it to be similar to the antitumor and antiviral drugs netropsin, distamycin, and the dye Hoechst 33258. Hoechst 33258 inhibits proper centromeric heterochromatin condensation and ultimately prevents formation of kinetochores in mouse chromosomes (Lica *et al.* 1986).

Alpha satellite DNA has long been considered a strong candidate for a functional human centromeric DNA sequence. Some key observations which support this postulation are: (1) Alpha satellite DNA is detectable at the centromere of all human chromosomes. (2) Alphoid monomers contain nucleotide sequences (e.g. CENP-B box) that are capable of specific binding to a number of nuclear proteins. (3) Preliminary functional studies appear to indicate that chromosome-integrated human alpha satellite DNA, exhibit some key features of a functioning centromere (Haaf *et al.* 1992).

Other studies appear to contradict the notion that functional human centromeres require alpha satellite DNA. (1) All alphoid DNA repeats isolated from human Y-chromosomes so far lack CENP-B boxes, an observation consistent with the inability to detect CENP-B proteins in the centromere of human and mouse Y-chromosomes. (2) A number of mitotically stable small marker chromosomes are being identified which appear to lack detectable amounts of alpha satellite DNA, despite the presence of an obviously functional centromere (Voullaire *et al.* 1993; Ohashi *et al.* 1994; Sacchi *et al.* 1996). Hopefully, a coherent resolution will soon be made to account for all of these seemingly contradictory observations.

3. Beta satellite DNA. The human Sau3A DNA family was identified by Meneveri *et al.* (1985) as a GC-rich repetitive DNA family comprised of 68 bp monomers. *In situ* hybridization experiments localized the 68 bp Sau3A clone (pUh1-39) to the pericentromeric regions of the human acrocentric chromosomes and the secondary constriction of chromosome 1 (Agresti *et al.* 1987, 1989).

Waye and Willard (1989a) proposed the name of β satellite DNA for this DNA family. They demonstrated that a 2.0 kb β satellite HOR (clone pB4) was primarily localized to the short arms of the acrocentric chromosomes. Homologous clones to pB4 (i.e. p21β2 and p21β7) were later used to define two distinct domains of beta satellite DNA, one domain proximal (p11) and one domain distal (p13) to the rRNA gene clusters in the short arm of all acrocentric chromosomes (Greig and Willard 1992). Another beta satellite clone (clone pB3 - 2.5 kb β satellite HOR) was mapped to human chromosome 9, first to the variable heterochromatin region of the long arm (i.e. 9qh) (Waye and Willard 1989a) and later to the centromeric region of this chromosome (Greig and Willard 1992). Cooper *et al.* (1992) also described approximately 54 kb of this DNA family in the centromeric region of the Y-chromosome. This communication suggested that these 68 bp repeats may actually be organized as a 16.9 kb HaeIII HOR in the Y-chromosome centromere.

4. Gamma satellite DNA. In 1992, Fan and coworkers performed physical mapping of cosmid clones from a flow sorted human X-chromosome library. Two clones were obtained which mapped unexpectedly to the centromeric region of human chromosome 8[13]. Chromosome 8-specific alphoid DNA failed to hybridize to restriction fragments from these two cosmid clones suggesting that these clones contained centromeric DNA sequences other than alpha satellite DNA (Fan *et al.* 1992).

Two EcoRI fragments, 50E1 (704 bp) and 50E4 (1962 bp), were obtained from one of the cosmid clones and mapped to the centromere of chromosome 8. Sequence analyses of these cloned DNA fragments revealed tandemly arranged, GC-rich repeat units of approximately 220 bp which lacked significant sequence similarity to any other previously identified human centromeric DNA sequence (Lin *et al.* 1993). This sequence was designated as a human gamma satellite DNA subfamily specific for chromosome 8 (i.e. gamma-8 satellite DNA). Both gamma-8 satellite DNA subclones were specifically localized to the primary constriction of chromosome 8, usually as two distinct fluorescent dots at the lateral sides of the centromere.

---

[13] It is presumed that some chromosome 8 material contaminated this chromosome X flow-sorted library because of the similar sizes of these two chromosomes.

5. 48 bp satellite DNA. During efforts to isolate polymorphic markers from a chromosome 22 specific DNA library, a 1.9 kb EcoRI fragment (clone p22hom48.4) was recovered from a lambda phage clone. Sequence data of a 382 bp PstI fragment from clone p22hom48.4 revealed tandemly arranged 48 bp repeats. Although this clone was initially localized to 22pter-q11, this DNA family is thought to reside in the centromeric region of chromosome 22 (Metzdorf *et al.* 1988).

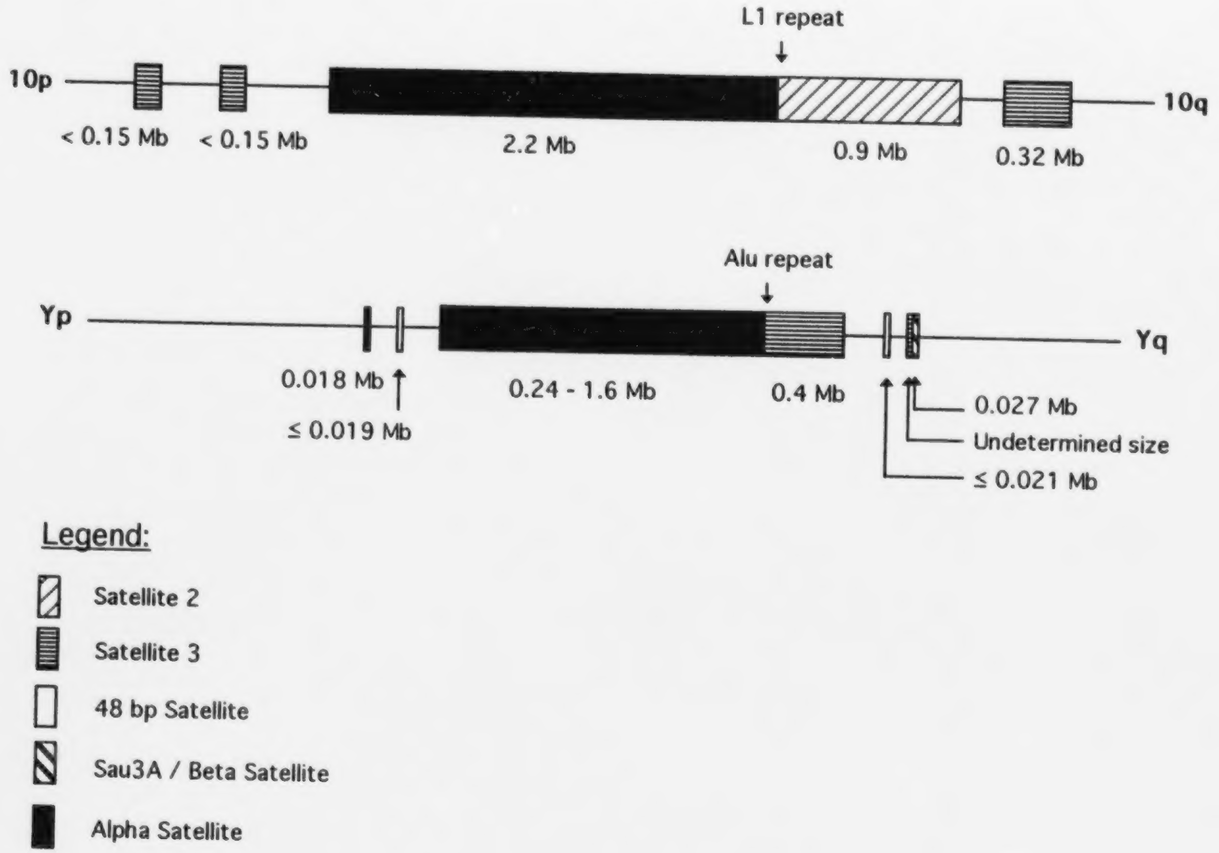
A second 48 bp subfamily was isolated from a chromosome 21-specific cosmid library (Mullenbach *et al.* 1992). Sequence data from a 507 bp clone (pcos6) revealed approximately 10 tandem repeats of 48 bp. The repeat units from chromosome 21 clones exhibited 4% to 31% sequence divergence from the 48 bp consensus sequence of chromosome 22. Low stringency *in situ* hybridization studies showed hybridization signals at the pericentromeric regions of chromosomes 13, 14, 15, and 21. Mullenbach *et al.* 1992 proposed a total array length of between 200 kb and 1100 kb for this DNA family in chromosome 21. The chromosome 21 DNA array(s) is/are approximately 340 kb away from alpha satellite DNA in this chromosome, as shown by pulsed field gel electrophoresis studies.

This DNA family was also identified in the centromeric region of the human Y-chromosome (Cooper *et al.* 1992, 1993). At least two 20 kb blocks of these 48 bp repeats were identified, flanking the major alphoid and satellite 3 DNA arrays (Figure 1.1). The 48 bp repeats in the chromosome Y centromere were approximately 88% identical to the published consensus sequence derived from homologous repeats in chromosome 22.

Restriction fragment length polymorphisms were detected in EcoRI- (Metzdorf *et al.* 1988), BamHI-, and MspI- (Mullenbach *et al.* 1992) digested genomic DNAs and gel mobility shift assays imply that two or more proteins form a complex with a single repeat unit of this DNA family (Mullenbach *et al.* 1992).

Figure 1.1. Genomic map of the relative size and distribution of identified repetitive DNA sequences in the centromeres of human chromosomes 10 and Y.

Spaces between DNA arrays in this figure are not to scale and do not reflect actual distances between blocks of centromeric repetitive DNAs. Adapted from Cooper *et al.* (1992, 1993) and Jackson *et al.* (1992, 1993).



L1 repeat
10p
10q
< 0.15 Mb
< 0.15 Mb
2.2 Mb
0.9 Mb
0.32 Mb
Alu repeat
Yp
Yq
0.018 Mb
≤ 0.019 Mb
0.24 - 1.6 Mb
0.4 Mb
0.027 Mb
Undetermined size
≤ 0.021 Mb
Legend:
Satellite 2
Satellite 3
48 bp Satellite
Sau3A / Beta Satellite
Alpha Satellite

6. Sn5 satellite DNA. Johnson *et al.* (1992) used chromosome microdissection and sequence independent PCR to construct a DNA library from marker chromosomes in peripheral lymphocytes of a new-born infant with a karyotype of 94,XXXX,+2mar. A 465 bp clone (pcrsn5) was isolated from this DNA library and localized to the pericentromeric regions of chromosomes 2, 20, and the acrocentric chromosomes. Homologous sequences were obtained from the African green monkey genome and found to share 69% sequence identity with the human pcrsn5 clone, over an aligned region of 241 bp (Johnson *et al.* 1992). From overlapping sequence comparisons of human and simian sequences, this DNA family was presumed to contain tandemly organized repeat units and was therefore designated as Sn5 satellite DNA. The monomer size of the human sn5 DNA family has not been determined but is believed to be greater than 406 bp. Sn5 satellite DNA is thought to have evolved in a similar manner as the alphoid suprachromosomal family 2, based on a similar chromosomal localization pattern (Table 1.3).

Interspersed repetitive DNA in human centromeres.

1. SINEs. Short interspersed repetitive elements (SINEs) are types of retroposons, which amplify themselves and become incorporated back into the human genome through transposition (reviewed in Rogers 1985, Okada 1991). The human Alu[14] SINE family (also referred to as SINE-1) consists of ~300 bp repeats which are reiterated every 3 - 6 kb throughout most of the human genome, except for chromosome regions like the centromere, where it is thought to be underrepresented some 50 fold (Moyzis *et al.* 1989). Alu repeats occasionally exist in human centromeres and have been found in or near alpha satellite DNA arrays (Jørgensen *et al.* 1986, Wevrick *et al.* 1992, Cooper *et al.* 1993,

---

[14] Alu sequences were first isolated by renaturing sheared (2 kb) denatured human placental DNA to a C_0t value of 68 and digestion of single stranded DNA with S_1 nuclease. Five percent of the products were 300 bp duplexes which when cleaved with the restriction enzyme AluI (hence the name of this repetitive DNA family), produced two characteristic fragments of approximately 170 bp and 120 bp (Houck *et al.* 1979). Ullu and Tschudi (1984) later established that Alu repeats were derived from 7SL cytoplasmic RNA, a component of the signal recognition particle.

Prades *et al.* 1996).

2. LINEs. Long interspersed repetitive elements (LINEs)[15] are transposable elements with repeat units of several kb in length (reviewed in Singer 1982, Hutchison *et al.* 1989). The mammalian Kpn LINE family (also known as LINE-1 or L1) has members as long as 6 - 7 kb in length[16] located every 30 - 60 kb in the human genome. Presumably, such LINE sequences are also substantially under represented in the human centromeres, accounting for their limited occurrences in human centromeric DNA arrays (e.g. Potter 1984, Wevrick *et al.* 1992, Prades *et al.* 1996).

Two studies have reported a single interspersed repetitive DNA element at the junction between an alpha satellite DNA array and a satellite 3 domain. Jackson *et al.* (1992) reported a 485 bp truncated and rearranged L1 element at a junction between these two DNA arrays on chromosome 10 (Figure 1.2). In the Y-chromosome centromere, an alphoid and satellite 3 junction harbors a single Alu repeat (Cooper *et al.* 1993). It is not clear why these retrotransposon elements would incorporate at these particular centromeric sites, although some investigators suggest that such sequences may have originated from molecular rearrangements within alpha satellite DNA arrays rather than from de novo insertions (Marçais *et al.* 1991, Prades *et al.* 1996).

3. 724 sequence family. A 724 sequence DNA family was identified by Kurnit *et al.* (1984) during a screening procedure for cDNA clones specific for chromosome 21. As a result of the screening procedure, a 0.6 kb clone, pUNC724, was isolated. Unfortunately, sequence data for this clone has not been published and the repeat unit length is undetermined. Based on an infrequent Southern blot hybridization to pools of recombinant human genomic DNA cosmids, it was suggested that this DNA family is interspersed in nature; although, this claim should be confirmed by direct sequence analyses. The location of this DNA family is still in dispute. The 724 DNA family has been reported in the pericentromeric regions of all human acrocentric chromosomes as well as at specific sites on the long arm of chromosomes 1, 18 (Kurnit *et al.* 1984), and 21 (Kurnit *et al.* 1986). Others suggest that the 724 DNA family resides primarily in the short arms of the acrocentric chromosomes (Vogt 1990, Choo *et al.* 1992). Copy number estimations originally suggested that the pUNC724 clone is repeated approximately 10 times in the human genome. However, this figure may be an underestimate, with the true copy number being on the order of 100 - 200 per genome (Vogt 1990).

---

[15] LINEs are readily distinguishable from SINEs, not only on the basis of repeat unit sizes, but also by the presence of two open reading frames (ORFs) within a LINE element. These ORFs potentially encode some or all enzymes required for retrotransposition.

[16] However, more than 90% of identified L1 repeats are actually truncated copies.

Unclassified human centromeric repetitive DNA sequences.

1. HaeIII DNA family. Some centromeric DNA repetitive families have yet to be categorized with certainty as satellite DNAs or interspersed repetitive DNAs. One such family is the HaeIII DNA family represented by a 291 bp fragment (clone A17e), isolated and sequenced by Agresti et al. (1989). Although an internal repetition was not shown in clone A17e, it was suggested that this DNA family consisted of repeat units which are 140 bp or 160 bp in length. Agresti *et al.* (1989) further suggested that this DNA family is in close proximity to some beta DNA arrays in the human genome.

2. ATRS DNA. Another unclassified human centromeric repeated sequence is a ~65% AT Rich Sequence (ATRS) discovered during the isolation of diverged chromosome 7 specific alpha satellite DNAs from a monochromosomal somatic cell hybrid (Wevrick *et al.* 1992). A phage clone, which hybridized at low but not high stringency conditions to two chromosome 7-specific alpha satellite DNA probes, contained both alpha satellite DNA and approximately 3.5 kb of the novel ATRS DNA. Analysis of 483 bp of ATRS failed to reveal any significant internal repetition making it difficult to conclude whether this sequence is tandemly organized or interspersed. FISH studies revealed hybridization signals primarily at or near the centromeres of all acrocentric chromosomes as well as chromosomes 1, 2, 7, 9, 16, and 17.

3. Centromeric repetitive elements from the Y-chromosome. During the construction of a long range genomic map of the pericentromeric region of the Y-chromosome, a number of putative repetitive elements were identified. Cooper *et al.* (1992) isolated three clones: pKFC37, pKFC43, and 64b, each of which yielded a unique pattern of multiple hybridizing restriction bands with total genomic DNA. Cooper *et al.* (1993) isolated three additional clones: YII5.1, YII1.1, and YII2.1, flanking a distinct 18 kb alphoid domain of highly diverged repeats in the Y centromere. YII5.1 seems to recognize a chromosome Y-specific low repeat whereas YII1.1 and YII2.1 represent portions of a moderately repetitive DNA sequence family which is also localized to other human chromosomes. Furthermore, 4 yeast artificial chromosomes (YACs), with human inserts from chromosomes other than the Y-chromosome, were isolated and each found to contain YII2.1 / YII1.1 homologous DNA sequences as well as alpha satellite DNA (Cooper *et al.* 1993). Based on this observation and copy number estimations of ~20 YII2.1 / YII1.1 per genome, it was suggested that at least one copy of YII2.1 and YII1.1 homologous sequences may reside in the centromeric region of most, if not all, human chromosomes. Genomic organization and sequence definition of these repetitive DNA families are anxiously awaited.

Cervid centromeric DNAs.

According to Whitehead (1993), there are currently 41 known deer species in the world. Among these 41 deer species, extreme chromosome number variation can be observed, ranging from 2n=80 in the Siberian roe deer (*Capreolus capreolus pygargus*) (Neitzel, 1987) to 2n=6 in the female Indian muntjac (*Mutiacus muntjak vaginalis*)[17]. Hence, this mammalian family is an excellent natural source for studies in karyotypic evolution.

---

[17] The Indian muntjac has the lowest chromosome of all known mammalian species.

Repetitive DNA sequences and the karyotypic evolution of the Asian muntjacs.

A longstanding enigma is based on the karyotypes of two Asian barking[18] deer species: the Indian muntjac (*Muntiacus muntjak vaginalis*) and the Chinese muntjac (*Muntiacus reevesi*). These two deer species are thought to be very closely related, based on similar phenotypic features and their ability to interbreed and produce viable F1 hybrids (Shi *et al.* 1980). However, their chromosome numbers vary remarkably. The Chinese muntjac has a diploid chromosome number of 46 (Wurster and Benirschke 1967b) and the Indian muntjac has a diploid chromosome number of 2n=6 (female) / 2n=7 (male)[19] (Wurster and Benirschke 1967a[20], 1970). All 46 chromosomes of the Chinese muntjac are acrocentric in morphology whereas the Indian muntjac karyotype is bimodal. Hsu *et al.* (1975) first proposed that the giant chromosomes of the Indian muntjac were derived by repeated centric and tandem fusions of smaller ancestral Chinese muntjac-like chromosomes. These authors speculated that chromosome breakages at the ends of the smaller acrocentric chromosomes (i.e. within the centromeric and/or telomeric heterochromatin) could have permitted the subsequent fusion of whole chromosome arms, leading to a drastically restructured karyotype with minimal loss or no loss of euchromatin (See Figure 1.3).

A number of investigations used chromosome banding techniques to test this "tandem chromosome fusion" theory in these Asian muntjacs. If chromosome breakage occasionally occurred at the short arm of the ancestral acrocentric chromosomes, interstitial remnants of centromeric heterochromatin would be expected in the Indian muntjac chromosomes. Comings (1971) thus employed a C-banding technique to identify such regions of interstitial centromeric heterochromatin. The results were inconclusive as no interstitial C-bands were observed along the arms of the Indian muntjac chromosomes. In 1980, Shi *et al.* examined the G-banded karyotype of a female hybrid produced from an Indian muntjac father and a Chinese muntjac mother. The 23 G-banded Chinese muntjac chromosomes in a metaphase spread of the hybrid were arranged in a specific fashion to imply that fusion of the Chinese muntjac chromosomes could have produced the specific G-banding pattern seen in the remaining three giant Indian muntjac chromosomes of the same metaphase spread.

---

[18] The term barking deer reflects the high pitched "barking" sound these deer make when sensing danger.

[19] In the Indian muntjac, the X chromosome is fused to the third autosome. Therefore, in XY males, the presence of only one X chromosome and the addition of a Y chromosome increases the total chromosome number in male Indian muntjacs by 1.

[20] Although the chromosome number of the female Indian muntjac (2n=6?) was first reported in the July 1967 issue of the Mammalian Chromosome Newsletter, the observation of few giant chromosomes was thought to be due to a culture contamination. Thus, the Indian muntjac karyotype was not confirmed until three years later. (Personal communications between Drs. Wurster-Hill and Lin.)

DNA studies by Wurster and Atkin (1972) revealed that the DNA content of the Indian muntjac cells is approximately 81% of the DNA content found in Chinese muntjac cells. Schmidtke *et al.* (1981) and Johnston *et al.* (1982) suggested that this difference in DNA content was attributed to the loss of a highly repetitive DNA fraction in the Indian muntjac genome. Bogenberger *et al.* (1985) and Yu *et al.* (1986) isolated centromeric satellite DNA sequences from the Indian muntjac genome and examined their distribution in both Indian and Chinese muntjac chromosomes. Radioactive *in situ* hybridization experiments suggested that these centromeric DNA sequences were confined to the centromeric regions of both the Chinese and Indian muntjac chromosomes and no hybridization signal was located at interstital sites of the Indian muntjac chromosomes. Lin *et al.* (1991) took a slightly different approach and isolated a highly repetitive centromeric DNA clone from the Chinese muntjac genome (clone C5) to examine its chromosomal distribution in the chromosomes of the Indian and Chinese muntjacs. In this case, fluorescent *in situ* hybridization results revealed non-random clusters of hybridization signals at the centromeric and at various interstitial sites of the Indian muntjac chromosomes. This suggested that remnants of centromeric heterochromatin did exist in the Indian muntjac chromosomes, possibly from repeated tandem chromosome fusions of ancestral Chinese muntjac-like chromosomes. These chromosome fusion events could have also left remnants of telomeric sequences at interstitial sites of the Indian muntjac chromosomes. Such interstitial sites were subsequently identified in the Indian muntjac chromosomes (Lee *et al.* 1993, Scherthan 1995) lending further evidence for the tandem chromosome fusion hypothesis.

Major cervid centromeric satellite DNA (cervid satellite I DNA).

As mentioned above, the first cervid centromeric DNA sequence to be cloned and sequenced was the Indian muntjac clone 1A (Bogenberger *et al.* 1985). This DNA clone represents a major cervid centromeric satellite DNA family[21], which shares sequence homology to bovine satellite I DNA[22], and is currently the only cervid centromeric satellite DNA family identified. Bogenberger *et al.* (1987) demonstrated by Southern blot analyses that this satellite DNA was highly conserved in the genomes of all deer species studied. More interestingly, this satellite DNA family was primarily organized into 0.8 kb monomer units in deer species belonging to the paleontological division, Plesiometacarpalia[23] and into 1 kb monomers in deer species belonging to the Telemetacarpalia division[24]. The presence of such distinct, different-sized monomers in a single satellite DNA family is highly unusual and raises questions regarding the genesis and evolution of this satellite DNA family.

When the first two cervid centromeric satellite DNA monomers were cloned from the Indian muntjac genome (Bogenberger *et al.* 1985, Yu *et al.* 1986), it was noted that both clones consisted of internal 31 bp tandem repeats (subrepeats). These 31 bp repeats shared significant sequence similarity to the 31 bp subrepeats found in bovine satellite I

---

[21] This satellite DNA family can be referred to as either major cervid centromeric satellite DNA or cervid satellite I DNA, because of its homology to bovine satellite DNA.

[22] This satellite DNA is sometimes referred to as bovine satellite 1.715 based on its bouyant density in CsCl gradients.

[23] Brooke (1878) divided all deer species into two main divisions. Those deer which have retained the more proximal remnants of the second and fifth metacarpals were classified into the plesiometacarpalia division whereas other deer were categorized into the telemetacarpalia division.

[24] With the exception of the European roe deer (*Capreolus capreolus*) which has its cervid satellite I DNA organized predominantly into 2 kb monomers (Scherthan 1991).

DNA (Plucienniczak *et al.* 1982). Subsequent reports of cervid satellite I DNA clones from the Chinese muntjac (clone C5 - Lin *et al.* 1991), the European roe deer (clone CC-SatI - Scherthan 1991), and the Canadian woodland caribou (clone Rt-Pst3 - Lee *et al.* 1994) have all failed to demonstrate the presence of any internal subrepeats.

With the idea that 31 bp subrepeats appeared to be present in the Indian muntjac centromeric satellite DNA clones, Bogenberger *et al.* (1987) suggested that the 1 kb monomers could have simply resulted from approximately 32 amplifications of the original 31 bp DNA sequence (as opposed to the approximately 26 amplification required to produce the 0.8 kb monomers). This postulation would predict that all subrepeats of a monomer were equally divergent from one another. Lee *et al.* (1994) cloned, sequenced, and characterized a 991 bp cervid satellite DNA monomer from the Canadian woodland caribou (clone Rt-Pst3) and found that the first ~800 bp of this clone to be similar to previously characterized 0.8 kb cervid monomers. The authors also found that the remaining 191 bp of the Rt-Pst3 clone actually shared 60% sequence similarity with the first ~191 bp of the same clone suggesting that perhaps some 1 kb cervid satellite I DNA monomers arose from a selective amplification of an ancestral 0.8 kb monomer along with ~191 bp of an adjacent monomer.

These hypotheses for the genesis of cervid satellite I DNA depended on sequence data from only two monomer clones from the Indian muntjac and a single monomer clone from the Canadian woodland caribou (*Rangifer tarandus caribou*). Clearly, more sequence data and analyses are required to test these postulations.

Rationale and scope of this thesis:

Despite the substantial number of investigations achieved on human centromeric satellite DNA families, DNA sequences essential for centromere function have yet to be determined. Recently, the discovery of the human gamma satellite DNA family prompts further characterization of this novel centromeric sequence.

Compared to humans, less is known of the centromeric satellite DNAs of other mammalian species. In the deer family, it has been proposed that centromeric satellite DNAs have contributed to the extreme karyotypic diversity observed. The unique karyotype of the Indian muntjac offers an excellent system for investigating the structure and function of mammalian centromeres. Furthermore, several questions have been raised regarding the genesis and evolution of the well conserved major cervid centromeric satellite DNA. Therefore, research has been carried out in this thesis to further characterize human gamma satellite DNA as well as address those questions regarding cervid centromeric satellite DNA.

Chapter 2 describes the isolation and characterization of a subfamily of human gamma satellite DNA specific for the centromeric region of the X-chromosome. Chapter 3 examines the presence of human gamma satellite DNA in two old world primate species as well as in a mitotically stable human marker chromosome. Chapter 4 details the identification of a 31 bp bovine subrepeat in cervid satellite I DNA clones from four different deer species. Finally, chapter 5 describes the higher-order organization of 31 bp subrepeats in cervid satellite I DNA family and postulates the genesis and evolution of this centromeric satellite DNA.

References:

Agresti A, Rainaldi G, Lobbiani A, Magnani I, Lernia RD, Meneveri R, Siccardi AG, Ginelli E. (1987) Chromosomal location by in situ hybridization of the human Sau3A family of DNA repeats. *Hum Genet* 75: 326-332.

Albertson DG, Thomson JN. (1982) The kinetochores of *Caenorhabditis elegans*. *Chromosoma* 86: 409-428.

Alexandrov IA, Mitkevich SP, Yurov YB. (1988) The phylogeny of human chromosome specific alpha satellites. *Chromosoma* 96: 443-453.

Alexandrov IA, Mashkova TD, Akopian TA, Medvedev LI, Kisselev LL, Mitkevich SP, Yurov YB. (1991) Chromosome-specific alpha satellites: Two distinct families on human chromosome 18. *Genomics* 11: 15-23.

Alexandrov IA, Medvedev LI, Mashkova TD, Kisselev LL, Romanova LY, Yurov YB. (1993) Definition of a new alpha satellite suprachromosomal family characterized by monomeric organization. *Nucleic Acids Res* 21: 2209-2215.

Baldini A, Rocchi M, Archidiacono N, Miller OJ, Miller DA. (1990) A human alpha satellite DNA subset specific for chromosome 12. *Am J Hum Genet* 46: 784-788.

Baldini A, Archidiacono N, Carbone R, Bolino A, Shridhar V, Miller OJ, Miller DA, Ward DC, Rocchi M. (1992) Isolation and comparative mapping of a human chromosome 20-specific α-satellite DNA clone. *Cytogenet Cell Genet* 59: 12-16.

Bayne RAL, Broccoli D, Taggart MH, Thomson EJ, Farr CJ, Cooke HJ. (1994) Sandwiching of a gene within 12 kb of a functional telomere and alpha satellite does not result in silencing. *Hum Mol Genet* 3: 539-546.

Bogenberger JM, Neumaier PS, Fittler F. (1985) The muntjak satellite IA sequence is composed of 31-base-pair internal repeats that are highly homologous to the 31-base-pair subrepeats of the bovine satellite 1.715. *Eur J Biochem* 148: 55-59.

Bokhari FS, Godward MBE. (1980) The ultrastructure of the diffuse kinetochore in *Luzula nivea*. *Chromosoma* 79: 125-136.

Britten RJ, Kohne DE. (1968) Repeated sequences in DNA. *Science* 161: 529-540.

Brooke V. (1878) On the classification of the Cervidae, with a synopsis of the existing species. *J Zool* 883-928.

Charlesworth B, Langely CH, Stephan W. (1986) The evolution of restricted recombination and the accumulation of repeated DNA sequences. *Genetics* 112: 947-962.

Choo KH, Brown R, Webb G, Craig IW, Filby RG. (1987) Genomic organization of human centromeric alpha satellite DNA: Characterization of a chromosome 17 alpha satellite sequence. *DNA* 6: 297-305.

Choo KH, Vissel B, Brown R, Filby RG, Earle E. (1988) Homologous alpha satellite sequences on human acrocentric chromosomes with selectivity for chromosome 13, 14, and 21: implications for recombination between nonhomologues and Robertsonian translocations. *Nucleic Acids Res* 16: 1273-1284.

Choo KH, Earle E, McQuillan C. (1990a) A homologous subfamily of satellite III DNA on human chromosomes 14 and 22. *Nucleic Acids Res* 18: 5641-5648.

Choo KH, Earle E, Vissel B, Filby RG. (1990b) Identification of two distinct subfamilies of alpha satellite DNA that are highly specific for human chromosome 15. *Genomics* 7: 143-151.

Choo KH, Vissel B, Nagy A, Earle E, Kalitsis P. (1991) A survey of the genomic distribution of alpha satellite DNA on all human chromosomes, and derivation of a new consensus sequence. *Nucleic Acids Res* 19: 1179-1182.

Choo KHA, Earle E, Vissel B, Kalitsis P. (1992) A chromosome 14-specific human satellite III DNA subfamily that shows variable presence on different chromosomes 14. *Am J Hum Genet* 50: 706-716.

Chuang CR, Saunders GF. (1974) Complexity of human satellite A DNA. *Biochem Biophys Res Commun* 57: 1221-1230.

Comings DE. (1971) Heterochromatin of Indian muntjac. *Exp Cell Res* 67: 441-460.

Cooke HJ, Hindley J. (1979) Cloning of human satellite III DNA: different components are on different chromosomes. *Nucleic Acids Res* 6: 3177-3197.

Cooke HJ, Schmidtke J, Gosden JR. (1982) Characterization of a human Y chromosome repeated sequence and related sequences in higher primates. *Chromosoma* 87: 491-502.

Cooke CA, Bernat RL, Earnshaw WC. (1990) CENP-B: A major human centromere protein located beneath the kinetochore. *J Cell Biol* 110: 1475-1488.

Cooper KF, Fisher RB, Tyler-Smith C. (1992) Structure of the pericentric long arm region of the human Y chromosome. *J Mol Biol* 228: 421-432.

Cooper KF, Fisher RB, Tyler-Smith C. (1993) Structure of the sequences adjacent to the centromeric alphoid satellite DNA array on the human Y chromosome. *J Mol Biol* 230: 787-799.

Corneo G, Ginelli E, Polli E. (1967) A satellite DNA isolated from human tissues. *J Mol Biol* 23: 619-622.

Corneo G, Ginelli E, Polli E. (1968) Isolation of the complementary strands of a human satellite DNA. *J Mol Biol* 33:331-335.

Corneo G, Ginelli E, Polli E. (1970) Repeated sequences in human DNA. *J Mol Biol* 48: 319-327.

Corneo G, Ginelli E, Polli E. (1971) Renaturation properties and localization in heterochromatin of human satellite DNAs. *Biochim Biophys Acta* 247: 528-534.

Corneo G, Zardi L, Polli E. (1972) Elution of human satellite DNAs on a methylated albumin kieselguhr chromatographic column: isolation of satellite DNA IV. *Biochim Biophys Acta* 269: 201-204.

D'Aiuto L, Antonacci R, Marzella R, Archidiacono N, Rocchi M. (1993) Cloning and comparative mapping of a human chromosome 4-specific alpha satellite DNA sequence. *Genomics* 18: 230-235.

Deininger PL, Jolly DJ, Rubin CM, Friedmann T, Schmid CW. (1981) Base sequence studies of 300 nucleotide renatured repeated human DNA clones. *J Mol Biol* 151:17-33.

Delattre O, Bernard A, Malfoy B, Marlhens F, Viegas-Pequignot E, Brossard C, Haguenauer O, Creau-Goldberg N, Van Cong N, Dutrillaux B, Thomas G. (1988) Studies on the human chromosome 3 centromere with newly cloned alphoid DNA probe. *Hum Hered* 38: 156-167.

Devilee P, Slagboom P, Cornelisse CJ, Pearson PL. (1986) Sequence heterogeneity within the human alphoid repetitive DNA family. *Nucleic Acids Res* 14: 2059-2073.

Devilee P, Kievits T, Waye JS, Pearson PL, Willard HF. (1988) Chromosome-specific alpha satellite DNA: Isolation and mapping of a polymorphic alphoid repeat from human chromosome 10. *Genomics* 3: 1-7.

Donlon TA, Bruns GA, Latt SA, Mulholland J, Wyman AR. (1987) A chromosome 8-enriched alphoid repeat. *Cytogenet Cell Genet* 46: 607.

Doolittle WF, Sapienza C. (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284: 601-603.

Durfy SJ, Willard HF. (1989) Patterns of intra and interarray sequence variation in alpha satellite from the human X chromosome: Evidence for short-range homogenization of tandemly repeated DNA sequences. *Genomics* 5: 810-821.

Earnshaw WC, Rothfield NF. (1985) Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma* 91: 313-321.

Earnshaw WC, Sullivan KF, Machlin PS, Cooke CA, Kaiser DA, Pollard TD, Rothfield NF, Cleveland DW. (1987) Molecular cloning of cDNA for CENP-B, the major human centromere autoantigen. *J Cell Biol* 104: 817-829.

Earnshaw WC, Ratrie H, Stetten G. (1989) Visualization of centromere proteins CENP-B and CENP-C on a stable dicentric chromosome in cytological spreads. *Chromosoma* 98: 1-12.

Fan YS, Sasi R, Lee C, Court D, Lin CC. (1992) Mapping of 50 cosmid clones isolated from a flow-sorted human X chromosome library by fluorescence *in situ* hybridization. *Genomics* 14: 542-545.

Frommer M, Prosser J, Tkachuk D, Reisner AH, Vincent PC. (1982) Simple repeated sequences in human satellite DNA. *Nucleic Acids Res* 10: 547-563.

Gaff C, du Sart D, Kalitsis P, Iannello R, Nagy A, Choo KHA. (1994) A novel nuclear protein binds centromeric alpha satellite DNA. *Hum Mol Genet* 3: 711-716.

Ge Y, Wagner MJ, Siciliano M, Wells DE. (1992) Sequence, higher order repeat structure, and long-range organization of alpha satellite DNA specific to human chromosome 8. *Genomics* 13: 585-593.

Gosden JR, Mitchell AR, Buckland RA, Clayton RP, Evans HJ. (1975) The location of four human satellite DNAs on human chromosomes. *Exp Cell Res* 92: 148-158.

Grady DL, Ratliff RL, Robinson DL, McCaniles EC, Meyne J, Moyzis RK. (1992) Highly conserved repetitive DNA sequences are present at human centromeres. *Proc Natl Acad Sci USA* 89: 1695-1699.

Greig GM, Willard HF. (1992) β satellite DNA: Characterization and localization of two subfamilies from the distal and proximal short arms of the human acrocentric chromosomes. *Genomics* 12: 573-580.

Greig GM, England SB, Bedford HM, Willard HF. (1989) Chromosome-specific alpha satellite DNA from the centromere of human chromosome 16. *Am J Hum Genet* 45: 862-872.

Greig GM, Parikh S, George J, Powers VE, Willard HF. (1991) Molecular cytogenetics of a α satellite DNA from chromosome 12: fluorescence in situ hybridization and description of DNA and array length polymorphisms. *Cytogenet Cell Genet* 56: 144-148.

Gruss P, Sauter G. (1975) Repetitive primate DNA containing the recognition sequences for two restriction endonucleases which generate cohesive ends. *FEBS Lett* 60: 85-88.

Haaf T, Willard HF. (1992) Organization, polymorphism, and molecular cytogenetics of chromosome-specific α-satellite DNA from the centromere of chromosome 2. *Genomics* 9: 122-128.

Haaf T, Warburton PE, Willard HF. (1992) Integration of human α-satellite into simian chromosomes: centromere binding and disruption of normal chromosome segregation. *Cell* 70: 681-696.

Haaf T, Steinlein K, Schmid M. (1986) Preferential somatic pairing between homologous heterochromatic regions of human chromosomes. *Am J Hum Genet* 38: 319-329.

Higgins MJ, Wang H, Shtromas I, Haliotis T, Roder JC, Holden JJA, White BN. (1985) Organization of a repetitive human 1.8 kb KpnI sequence localized in the heterochromatin of chromosome 15. *Chromosoma* 93: 77-86.

Houck CM, Rinehart FP, Schmid CW. (1979) A ubiquitous family of repeated DNA sequences in the human genome. *J Mol Biol* 132: 289-306.

Hsu TC, Pathak S, Chen TR. (1975) The possibility of latent centromeres and a proposed nomenclature system for total chromosome and whole arm translocations. *Cytogenet Cell Genet* 15: 41-49.

Hughes-Schrader S, Ris H. (1941) The diffuse spindle attachment of cocids, verified by the mitotic behavior of induced chromosome fragments. *J Exp Zool* 87: 429-451.

Hulsebos T, Schonk D, van Dalen I, Coerwinkel-Driessen M, Schepens J, Ropers HH, Wieringa B. (1988) Isolation and characterization of alphoid DNA sequences specific for the pericentric regions of chromosomes 4, 5, 9, and 19. *Cytogenet Cell Genet* 47: 144-148.

Hutchison CAH III, Hardies SC, Loeb DD, Shehee WR, Edgell MH. (1989) LINEs and related retroposons: Long interspersed repeated sequences in the eukaryotic genome. In: Mobile DNA. Berg DF, Howe MM, eds. Washington DC, American Society for Microbiology, pp. 593-617.

Jabs EW, Carpenter N. (1988) Molecular cytogenetic evidence for amplification of chromosome-specific alphoid sequences at enlarged C-bands on chromosome 6. *Am J Hum Genet* 43: 69-74.

Jabs EW, Persico MG. (1987) Characterization of human centromeric regions of specific chromosomes by means of alphoid DNA sequences. *Am J Hum Genet* 41: 374-390.

Jabs EW, Wolf SF, Migeon BR. (1984) Characterization of a cloned DNA sequence that is present at centromeres of all human autosomes and the X chromosome and shows polymorphic variation. *Proc Natl Acad Sci USA* 81: 4884-4888.

Jabs EW, Goble CA, Cutting GR. (1989) Macromolecular organization of human centromeric regions reveals high-frequency, polymorphic macro DNA repeats. *Proc Natl Acad Sci USA* 86: 202-206.

Jackson MS, Mole SE, Ponder BA. (1992) Characterization of a boundary between satellite III and alphoid sequences on human chromosome 10. *Nucleic Acids Res* 20: 4781-4787.

Jackson MS, Slijepcevic P, Ponder BAJ. (1993) The organization of repetitive sequences in the pericentromeric region of human chromosome 10. *Nucleic Acids Res* 21: 5865-5874.

Jeanpierre M. (1994) Human satellites 2 and 3. *Annales de Genetique* 37: 163-171.

Johnson DH, Kroisel PM, Klapper HJ, Rosenkranz W. (1992) Microdissection of a human marker chromosome reveals its origin and a new family of centromeric repetitive DNA. *Hum Mol Genet* 1: 741-747.

Johnston FP, Church RB, Lin CC. (1982) Chromosome rearrangement between the Indian and Chinese muntjac is accompanied by a deletion of middle repetitive DNA. *Can J Biochem* 60: 497-506.

Jones KW, Prosser J, Corneo G, Ginelli E. (1973) The chromosomal location of human satellite DNA III. *Chromosoma* 42: 445-451.

Jones KW, Purdom IF, Prosser J, Corneo G. (1974) The chromosomal localisation of human satellite DNA I. *Chromosoma* 49: 161-171.

Jørgensen AL, Bostock CJ, Bak AL. (1986) Chromosome-specific subfamilies within human alphoid repetitive DNA. *J Mol Biol* 187: 185-196.

Kalitsis P, Earle E, Vissel B, Shaffer LG, Choo KHA. (1993) A chromosome 13-specific human satellite I DNA subfamily with minor presence on chromosome 21: further studies on Robertsonian translocations. *Genomics* 16: 104-112.

Karpen GH, Le M-H, Le H. (1996) Centric heterochromatin and the efficiency of achiasmate disjunction in *Drosophila* female meiosis. *Science* 273: 118-122.

Kit S. (1961) Equilibrium centrifugation in density gradients of DNA preparations from animal tissue. *J Mol Biol* 3: 711-716.

Kurnit DM, Neve Rl, Marton CC, Bruns GAP, Ma NSF, Cox DR, Klinger HP. (1984) Recent evolution of DNA sequence homology in the pericentromeric regions of human acrocentric chromosomes. *Cytogenet Cell Genet* 38: 99-105.

Kurnit DM, Roy S, Stewart GD, Schwedock J, Neve RL, Bruns GA, Van Keuren P, Patterson D. (1986) The 724 family of DNA sequences is interspersed about the pericentromeric regions of human acrocentric chromosomes. *Cytogenet Cell Genet* 43: 109-116.

Larin Z, Fricker MD, Tyler-Smith C. (1994) *De novo* formation of several features of a centromere following introduction of a Y alphoid YAC into mammalian cells. *Hum Mol Genet* 3: 689-695.

Lee C, Sasi R, Lin CC. (1993) Interstitial localization of telomeric DNA sequences in the Indian muntjac chromosomes: further evidence for tandem chromosome fusions in the karyotypic evolution of the Asian muntjacs. *Cytogenet Cell Genet* 63: 156-159.

Lee C, Ritchie DBC, Lin CC. (1994) A tandemly repetitive, centromeric DNA sequence from the Canadian woodland caribou (*Rangifer tarandus caribou*): its conservation and evolution in several deer species. *Chromosome Res* 2: 293-306.

Lica LM, Narayanswami S, Hamkalo BA. (1986) Mouse satellite DNA, centromere structure, and sister chromatid pairing. *J Cell Biol* 103: 1145-1151.

Lin CC, Sasi R, Fan Y-S, Chen Z-Q. (1991) New evidence for tandem chromosome fusions in the karyotypic evolution of Asian muntjacs. *Chromosoma* 101: 19-24.

Lin CC, Sasi R, Lee C, Fan YS, Court D. (1993) Isolation and identification of a novel tandemly repeated DNA sequence in the centromeric region of human chromosome 8. *Chromosoma* 102: 333-339.

Lund T, Holtlund J, Laland SG. (1985) On the phosphorylation of low molecular mass HMG (high mobility group) proteins in Ehrlich ascites cells. *FEBS Lett* 180: 275-279.

Maio JJ. (1971) DNA strand reassociation and polyribonucleotide binding in the African Green monkey, *Cercopithecus aethiops*. *J Mol Biol* 56: 579-595.

Mahtani MM, Willard HF. (1990) Pulsed-field gel analysis of α-satellite DNA at the human X chromosome centromere: High-frequency polymorphisms and array size estimate. *Genomics* 7: 607-613.

Manuelidis L, Wu JC. (1978) Homology between human and simian repeated DNA. *Nature* 276: 92-94.

Marçais B, Charlieu JP, Allain B, Brun E, Bellis M, Roizès G. (1991) On the mode of evolution of alpha satellite DNA in human populations. *J Mol Evol* 33: 42-48.

Mashkova TD, Akopian TA, Romanova LY, Mitkevich SP, Yurov YB, Kisselev LL, Alexandrov IA. (1994) Genomic organization, sequence and polymorphism of the human chromosome 4-specific α-satellite DNA. *Gene* 140: 211-217.

Masumoto H, Masukata H, Muro Y, Nozaki N, Okazaki T. (1989) A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *J Cell Biol* 109: 1963-1973.

McDermid HE, Duncan AMV, Higgins MJ, Hamerton JL, Rector E, Brasch KR, White BN. (1986) Isolation and characterization of an α-satellite repeated sequence from human chromosome 22. *Chromosoma* 94: 228-234.

Meneveri R, Agresti A, Della Valle G, Talarico D, Siccardi AG, Ginelli E. (1985) Identification of a human clustered G+C-rich DNA family of repeats (Sau3A Family).

J Mol Biol 186: 483-489.

Messelson M, Stahl FW, Vinograd. (1957) Equilibrium sedimentation of macromolecules in density gradients. *Proc Natl Acad Sci USA* 43: 581-588.

Metzdorf R, Göttert E, Blin N. (1988) A novel centromeric repetitive DNA from human chromosome 22. *Chromosoma* 97: 154-158.

Meyne J, Goodwin EH, Moyzis RK. (1994) Chromosome localization and orientation of the simple sequence repeat of human satellite I DNA. *Chromosoma* 103:99-103.

Mitchell AR, Beauchamp RS, Bostock CJ. (1979) A study of sequence homologies in four satellite DNAs of man. *J Mol Biol* 135: 127-149.

Mitchell AR, Gosden JR, Miller DA. (1985) p82H: a cloned sequence of the alphoid repeated DNA family found at the centromeres of all human chromosomes. *Chromosoma* 92: 369-377.

Moyzis RK, Albright KL, Bartholdi MF, Cram LS, Deaven LL, Hildebrand CE, Joste NE, Longmire JL, Schwarzacher-Robinson T. (1987) Human chromosome specific repetitive DNA sequences : Novel markers for genetic analysis. *Chromosoma* 95: 375-386.

Moyzis RK, Torney DC, Meyne J, Buckingham JD, Wu JR, Burks C, Sirotkin KM, Goad WB. (1989) The distribution of interspersed repetitive DNA sequences in the human genome. *Genomics* 4: 273-289.

Mullenbach R, Lutz S, Holzmann K, Dooley S, Blin N. (1992) A non-alphoid repetitive DNA sequence from human chromosome 21. *Hum Genet* 89: 519-523.

Muro Y, Masumoto H, Yoda K, Nozaki N, Ohashi M, Okazaki T. (1992) Centromere protein B assembles human centromeric alpha satellite DNA at the 17-bp sequence, CENP-B box. *J Cell Biol* 116: 585-596.

Nakahori Y, Mitani N, Yamada M, Nakagome Y. (1986) A human Y-specific repeated DNA family (DYZ1) consists of a tandem array of pentanucleotides. *Nucleic Acids Res* 14: 7569-7580.

Ohashi H, Wakui K, Ogawa K, Okano T, Tiikawa N, Fukushima Y. (1994) A stable marker chromosome: possible existence of an intercalary ancient centromere at distal 8p. *Am J Hum Genet* 55: 1202-1208.

Okada N. (1991) SINEs. *Curr Opin Genet Dev* 1: 498-504.

Orgel LE, Crick FHC. (1980) Selfish genes: the ultimate parasite. *Nature* 284: 604-607.

Ouspenski LL, Brinkley BR. (1993) Centromeric DNA cloned from functional kinetochore fragments in mitotic cells with unreplicated genomes. *J Cell Science* 105: 359-367.

Palmer DK, O'Day K, Trong HL, Charbonneau H, Margolis RL. (1991) Purification of the centromeric protein CENP-A and demonstration that it is a centromere specific histone. *Proc Natl Acad Sci USA* 88: 3734-3738.

Pech M, Igo-Kemenes T, Zachau HG. (1979) Nucleotide sequence of a highly repetitive component of rat DNA. *Nucleic Acids Res* 7: 417-432.

Plucienniczak A, Skowronski J, Jaworski J. (1982) Nucleotide sequence of bovine 1.715 satellite DNA and its relation to other bovine satellite sequences. *J Mol Biol* 158: 293-304.

Pluta AF, Cooke CA, Earnshaw WC. (1990) Structure of the human centromere at metaphase. *Trends Biochem Sci* 15: 181-185.

Potter SS. (1984) Rearranged sequences of a human KpnI element. *Proc Natl Acad Sci USA* 81: 1012-1016.

Prades C, Laurent A-M, Puechberty J, Yurov Y, Roizès G. (1996) SINE and LINE within human centromeres. *J Mol Evol* 42: 37-43.

Prosser J, Reisner AH, Bradley ML, Ho K, Vincent PC. (1981) Buoyant density and hybridization analysis of human DNA sequences, including three satellite DNAs. *Biochim Biophys Acta* 656: 93-102.

Prosser J, Frommer M, Paul C, Vincent PC. (1986) Sequence relationships of three human satellite DNAs. *J Mol Biol* 187: 145-155.

Reeves R, Nissen MS. (1990) The A·T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. *J Biol Chem* 265: 8573-8582.

Rocchi M, Baldini A, Archidiacono N, Lainwala S, Miller OJ, Miller DA. (1990) Chromosome-specific subsets of human alphoid DNA identified by a chromosome 2-derived clone. *Genomics* 8: 705-709.

Rocchi M, Archidiacono N, Ward DC, Baldini A. (1991) A human chromosome 9-specific alphoid DNA repeat spatially resolvable from satellite 3 DNA by fluorescent in situ hybridization. *Genomics* 9: 517-523.

Rogers J. (1985) The structure and evolution of retroposons. *Int Rev Cytol* 93: 187-279.

Rosenberg M, Segal S, Kuff EL, Singer MF. (1977) The nucleotide sequence of repetitive monkey DNA found in defective simian virus 40. *Cell* 11: 845-857.

Rosenberg H, Singer MF, Rosenberg M. (1978) Highly reiterated sequences of SIMIAN SIMIANSIMIANSIMIANSIMIAN. *Science* 200: 394-402.

Sacchi N, Magnani I, Fuhrman-Conti AM, Monard SP, Darfler M. (1996) A stable marker chromosome with a cryptic centromere: evidence for centromeric sequences associated with an inverted duplication. *Cytogenet Cell Genet* 73: 123-129.

Saunders GF, Hsu TC, Getz MJ, Simes EL, Arrighi FE. (1972) Locations of a human satellite DNA in human chromosomes. *Nature New Biology* 236: 244-246.

Scherthan H. (1991) Characterization of a tandem repetitive sequence cloned from the deer *Capreolus capreolus* and its chromosomal localisation in two muntjac species. *Hereditas* 115: 43-49.

Scherthan H. (1995) *Chromosome evolution in muntjac revealed by centromere, telomere and whole chromosome paint probes*. In: Brandham PE, Bennett MD, eds. Kew Chromosome Conference IV. Kew, Royal Botanic Gardens. pp. 267-280.

Schmidtke J, Brennecke H, Schmid H, Neitzel H, Sperling K. (1981) Evolution of Muntjac DNA. *Chromosoma* 84: 187-193.

Schulman I, Bloom KS. (1991) Centromeres: An integrated protein /DNA complex required for chromosome movement. *Ann Rev Cell Biol* 7: 311-336.

Schwarzacher-Robinson T, Cram LS, Meyne J, Moyzis RK. (1988) Characterization of human heterochromatin by in situ hybridization with satellite DNA clones. *Cytogenet Cell Genet* 47: 192-196.

Shaul Y, Garcia PD, Schonberg S, Rutter WJ. (1986) Integration of hepatitis B virus DNA in chromosome-specific satellite sequences. *J Virology* 59: 731-734.

Shi LM, Ye YY, Duan XS. (1980) Comparative cytogenetic studies on the red muntjac, Chinese muntjac and their F_1 hybrids. *Cytogenet Cell Genet* 26: 22-27.

Singer MF. (1982) Highly repeated sequences in mammalian genomes. *Int Rev Cytol* 76: 67-112.

Smith GP. (1976) Evolution of repeated DNA sequences by unequal crossover. *Science* 191: 528-535.

Solomon MJ, Strauss F, Varshavsky A. (1986) A mammalian high mobility group protein recognizes any stretch of six A·T base pairs in duplex DNA. *Proc Natl Acad Sci USA* 83: 1276-1280.

Strauss F, Varshavsky A. (1984) A protein binds to a satellite DNA repeat at three specific sites that would be brought into mutual proximity by DNA folding in the nucleosome. *Cell* 37: 889-901.

Sullivan KF, Hechenberger M, Masri K. (1994) Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. *J Cell Biol* 127: 581-592.

Tagarro I, Ferndández-Peralta AM, González-Aguilera JJ. (1994a) Chromosomal localization of human satellites 2 and 3 by FISH method using oligonucleotides as probes. *Hum Genet* 93: 383-388.

Tagarro I, Wiegant J, Raap AK, González-Aguilera JJ, Fernández-Peralta AM. (1994b) Assignment of human satellite 1 DNA as revealed by fluorescent in situ hybridization with oligonucleotides. *Hum Genet* 93: 125-128.

Tyler-Smith C, Brown WRA. (1987) Structure of the major block of alphoid satellite DNA on the human Y chromosome. *J Mol Biol* 195: 457-470.

Ullu E, Tschudi C. (1984) Alu sequences are processed 7SL RNA genes. *Nature* 312: 171-172.

Vissel B, Nagy A, Choo KHA. (1992) A satellite III sequence shared by human chromosomes 13, 14, and 21 that is contiguous with α satellite DNA. *Cytogenet Cell Genet* 61: 81-86.

Vogt P. (1990) Potential genetic functions of tandem repeated DNA sequence blocks in the human genome are based on a highly conserved "chromatin folding code". *Hum Genet* 84: 301-336.

Voullaire LE, Slater HR, Petrovic V, Choo KHA. (1993) A functional marker centromere with no detectable alpha-satellite, satellite III, or CENP-B protein: activation of a latent centromere? *Am J Hum Genet* 52: 1153-1163.

Wang HS, Riddell DC, Donald LJ, Cameron EC, Tonogai J, Holden JJA, Higgins MJ, Shtromas I, White BN, Hamerton JL. (1984) Mapping of a repetitive 1.8 kb Kpn 1 sequence to the short arm of chromosome 15. Human gene mapping 7. *Cytogenet Cell Genet* 37: 601-602.

Waring M, Britten RJ. (1966) Nucleotide sequence repetition: a rapidly reassociating fraction of mouse DNA. *Science* 154: 791-794.

Waye JS, Willard HF. (1986) Structure, organization, and sequence of alpha satellite DNA from human chromosome 17: Evidence for evolution by unequal crossing-over and an ancestral pentamer repeat shared with the human X chromosome. *Mol Cell Biol* 6: 3156-3165.

Waye JS, Willard HF. (1989a) Human β satellite DNA: Genomic organization and sequence definition of a class of highly repetitive tandem DNA. *Proc Natl Acad Sci USA* 86: 6250-6254.

Waye JS, Willard HF. (1989b) Chromosome specificity of satellite DNAs: short- and long-range organization of a diverged dimeric subset of human alpha satellite from chromosome 3. *Chromosoma* 97: 475-480.

Waye JS, Creeper LA, Willard HF. (1987a) Organization and evolution of alpha satellite DNA from human chromosome 11. *Chromosoma* 95: 182-188.

Waye JS, England SB, Willard HF. (1987b) Genomic organization of alpha satellite on human chromosome 7: evidence for two distinct domains on a single chromosome. *Mol Cell Biol* 7: 349-356.

Waye JS, Durfy SJ, Pinkel D, Kenwrick S, Patterson M, Davies KE, Willard HF. (1987c) Chromosome-specific alpha satellite DNA from human chromosome 1: Hierarchical structure and genomic organization of a polymorphic domain spanning several hundred kilobase pairs of centromeric DNA. *Genomics* 1: 43-51.

Waye JS, Mitchell AR, Willard HF. (1988) Organization and genomic distribution of "82H" alpha satellite DNA: evidence for a low-copy or single-copy domain on human chromosome 14. *Hum Genet* 78: 27-32.

Wevrick R, Willard HF. (1989) Long-range organization of tandem arrays of alpha satellite DNA at the centromeres of human chromosomes: high-frequency array-length polymorphism and meiotic stability. *Proc Natl Acad Sci USA* 86: 9394-9398.

Wevrick R, Willard VP, Willard HF. (1992) Structure of DNA near long tandem arrays of alpha satellite DNA at the centromere of human chromosome 7. *Genomics* 14: 912-923.

Widom J, Klug A. (1985) Structure of the 300 Å chromatin filament: X-ray diffraction from oriented samples. *Cell* 43: 207-213.

Willard HF, Waye JS. (1987a) Chromosome-specific subsets of human alpha satellite DNA: analysis of sequence divergence within and between chromosomal subsets and evidence for an ancestral pentameric repeat. *J Mol Evol* 25: 207-217.

Willard HF, Waye JS. (1987b) Hierarchical order in chromosome-specific human alpha satellite DNA. *Trends in Genet* 3: 192-198.

Wolfe J, Darling SM, Erickson RP, Craig IW, Buckle VJ, Rigby PWJ, Willard HF, Goodfellow PN. (1985) Isolation and characterization of an alphoid centromeric repeat family from the human Y chromosome. *J Mol Biol* 182: 477-485.

Wong AKC, Rattner JB. (1988) Sequence organization and cytological localization of the minor satellite of mouse. *Nucleic Acids Res* 16:11645-11661.

Wu JC, Manuelidis L. (1980) Sequence definition and organization of a human repeated DNA. *J Mol Biol* 142: 363-386.

Wurster DH, Atkin NB. (1972) Muntjac chromosomes: a new karyotype for *Muntiacus muntjac*. *Experimentia* 28: 972-973.

Wurster DH, Benirschke K. (1967a) The chromosomes of twenty three species of the Cervoidea and Bovidea. *Mammalian Chromosome Newsletter* 8:226-229.

Wurster DH, Benirschke K. (1967b) Chromosome studies in some deer, the springbok and the pronghorn, with notes on placentation in deer. *Cytologia* 32: 273-285.

Wurster DH, Benirschke K. (1970) Indian muntjac, *Muntiacus muntjak*: A deer with a low diploid chromosome number. *Science* 168: 1364-1366.

Yang TP, Hansen SK, Oishi KK, Ryder OA, Hamkalo BA. (1982) Characterization of a cloned repetitive DNA sequence concentrated on the human X chromosome. *Proc Natl Acad Sci USA* 79: 6593-6597.

Yoda K, Kitagawa K, Masumoto H, Muro Y, Okazaki T. (1993) A human centromere protein, CENP-B, has a DNA binding domain containing four potential α helices at the NH_2 terminus, which is separable from dimerizing activity. *J Cell Biol* 119: 1413-1427.

Yu LC, Lowensteiner D, Wong EFK, Sawada I, Mazrimas J, Schmid C. (1986) Localization and characterization of recombinant DNA clones derived from the highly repetitive DNA sequences in the Indian muntjac cells: Their presence in the Chinese muntjac. *Chromosoma* 93: 521-528.

CHAPTER 2

DISCOVERY AND PRELIMINARY CHARACTERIZATION OF HUMAN GAMMA X SATELLITE DNA

A version of this chapter has been published:

Lee C, Li X, Jabs EW, Court DR, Lin CC. (1995) Human gamma X satellite DNA: An X chromosome specific centromeric DNA sequence. *Chromosoma* 104: 103-112.

Introduction:

Centromeric DNA sequences which are responsible for proper chromosome segregation have been defined in two species of yeast, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* (reviewed in Clarke 1990). However, such functional centromeric DNA sequences have yet to be discovered in other eukaryotic species. Among higher eukaryote species, human centromeric DNA sequences have been relatively well characterized. Of all known human centromeric DNAs, alpha satellite DNA (Manuelidis 1978) is predominant and was once even considered to be the only centromeric DNA contributing to the structure and function of human centromeres (Willard 1990). This belief could stem from the following observations. First, alpha satellite has been found in the centromeric regions of all human chromosomes (Mitchell *et al.* 1985). Second, monomers of some alpha satellite DNA subsets were shown to harbor a 17 bp motif which serves as a binding site for the centromeric protein, CENP-B (Masumoto *et al.* 1989; Ikeno *et al.* 1994). Third, transfection studies with African Green monkey chromosomes (Haaf *et al.* 1992) and telomere directed fragmentation experiments which dissect the centromere of the human Y-chromosome (Brown *et al.* 1994) both suggest that alpha satellite DNA may be all that is necessary for proper chromosome segregation. On the other hand, a number of studies have shown that monospecific antibodies to the CENP-B protein recognize both active and inactive centromeres of dicentric chromosomes (Earnshaw *et al.* 1989; Page *et al.* 1995). Also, certain marker chromosomes, which are mitotically stable and thus thought to possess functional centromeres, appeared to lack alpha satellite DNA (Voullaire *et al.* 1993; Ohashi *et al.* 1994). Cumulatively, these observations lend to the continued debate on how essential alpha satellite DNA is in the formation of a functional human centromere.

In addition to alpha satellite DNA, other satellite DNAs have also been isolated and localized to the centromeric regions of human chromosomes. For example, the classical satellite DNAs have been detected in the centromeric regions of human chromosomes 3 (sat

I), 4 (sat I), 9 (sat 2, III), 13 (sat I, III), 14 (sat I, III), 15 (sat I, 2), 21 (sat I, III), 22 (sat I), and the Y-chromosome (sat 2) (Gosden *et al.* 1975; Higgins *et al.* 1985; Vissel *et al.* 1992; Meyne *et al.* 1994). Grady *et al.* (1992) suggested that the major repetitive component (i.e. GGAAT) of human classical satellites 2 and III could be present in the centromeres of almost every human chromosome. A subset of beta satellite DNA has been reported in the centromeric region of human chromosome 9 (Waye and Willard 1989). The centromeric region of the human acrocentric chromosomes appear to harbor a 406 bp tandem DNA repeat (Johnson *et al.* 1992), and Blin and co-workers reported a 48 bp tandemly repetitive element in the pericentromeric region of chromosome 21 (Mullenbach *et al.* 1992) and chromosome 22 (Metzdorf *et al.* 1988). In addition to satellite DNAs, Kurnit *et al.* (1984, 1986) described an interspersed 724 repetitive DNA in the pericentromeric regions of all human acrocentric chromosomes and Wevrick *et al.* (1992) described the presence of L1 repeats, an Alu element, and a novel AT rich repeat sequence (ATRS) in the centromeric region of human chromosome 7. Homologous ATRS sequences were also demonstrated in the pericentromeric regions of chromosomes 1, 2, and 16 (Wevrick *et al.* 1992). Thus, the human centromere is more heterogeneous, with respect to its DNA composition, than was previously thought (Willard 1992).

To this list of repetitive DNAs, a new satellite DNA sequence was recently identified in the centromeric region of human chromosome 8 and named gamma 8 satellite DNA (Lin *et al.* 1993). Gamma 8 satellite DNA is comprised of 220 bp tandemly organized repeats, which localize specifically to the centromere of human chromosome 8, often as two distinct dots at the lateral sides of the primary constriction. This could suggest a close proximity of gamma 8 satellite DNA to the kinetochore, implying a structural and/or functional role for this centromeric DNA. If gamma satellite DNA indeed has some essential role in the function of human centromeres, one would expect gamma satellite DNA to be present in the centromeres of other human chromosomes. Here, we report the isolation and preliminary characterization of a subfamily of gamma satellite DNA in the centromeric region of the human X-chromosome.

Materials and Methods:

Isolation of the 2D12/E2 subclone. A cosmid clone, CX16-2D12, containing approximately 40 kb of human genomic DNA was specifically localized to the centromeric region of the X-chromosome by fluorescence *in situ* hybridization (FISH). Chromosome X-specific alpha satellite DNA did not hybridize to any BamHI fragments of CX16-2D12 DNA suggesting that the cosmid clone lacked alpha satellite DNA (Fan *et al.* 1992). To investigate the nature of this centromeric DNA, a 1.2 kb EcoRI-digested fragment of CX16-2D12 was subcloned into pUC 18 following standard cloning protocols (Maniatis *et al.* 1982) and designated as 2D12/E2.

Southern blot hybridizations. Ten microgram aliquots of genomic DNA from nine unrelated individuals (7 females and 2 males) were digested separately with EcoRI, PstI, or SstI, and electrophoresed through a 0.8% (w/v) agarose gel. Fractionated genomic DNAs were transferred onto GeneScreen Plus nylon membranes (New England Nuclear) and hybridized to ^{32}P-dCTP labeled 2D12/E2 insert DNA at 58°C for 16 h in a hybridization mixture containing 2x SSC, 10x Denhardt's, 10% (w/v) dextran sulfate, 0.4% (w/v) SDS, 100 mg/ml sheared sonicated salmon sperm DNA, and 100 mg/ml yeast tRNA. Membranes were washed to a stringency of 0.1x SSC (15 mM NaCl, 1.5 mM sodium citrate) / 0.4% (w/v) SDS at 60 °C.

Ten millilitre blood samples were obtained from each of 8 unrelated individuals (6 males, 2 females) from which leukocytes were isolated with equal volumes of Ficoll. Isolated leukocytes from each individual were embedded in 1 ml of 1% (w/v) low melting point agarose (electrophoresis grade, BRL) after a single wash with PBS. The leukocytes embedded in each agarose block were then digested in an NDS / proteinase K solution (0.5

M EDTA, pH=9.5; 1% (w/v) N-lauroylsarcosine, 2 mg/ml proteinase K) for 24 h at 50°C, until the slightly opaque blocks became clear. EDTA and proteinase K were removed by one 12 h dialysis against low TE supplemented with 0.1 M phenylmethyl sulfonyl fluoride (in isopropanol) and one 12 h dialysis against low TE alone. Fifteen units of restriction enzyme were then added to each block of genomic DNA. The blocks were left on ice for 2 hours after which acetylated BSA (final concentration = 200 mg / ml) and another fifteen units of restriction enzyme were added to each block. Restriction digestions proceeded for 16 h in a 37 °C water bath and the digested products were electrophoresed through a 1% agarose gel for 16 h at 150 V with a 10 sec switch time using a BRL Hex-A-Field horizontal gel electrophoresis apparatus. Southern blot transfers of pulsed-field gels to nylon membranes, hybridization to ^{32}P-labeled 2D12/E2 probe DNA, and washings were performed as with the blots of DNAs fractionated by conventional gels.

DNA sequencing. The 2D12/E2 subclone was sequenced with a Sequenase kit (United States Biochemical Corporation) which employs the dideoxy-chain termination method. Sequence data were analyzed on the Sequence Editor program (Applied Biosystems) for Macintosh computers and the complete DNA sequence of 2D12/E2 was deposited into the EMBL database library (Accession number: X87951).

Copy number determination. Serial dilutions of EcoRI-digested 2D12/E2 plasmid DNA (0.2 mg - 0.10 ng) and sheared human female genomic DNA (4 mg - 0.06 mg) were loaded into separate compartments of a slot blot apparatus (Tyler Research) after the total DNA concentration for each dilution was adjusted to 4 mg with sheared salmon sperm DNA. The DNA samples were vacuum blotted onto a Gene Screen Plus membrane and hybridized to ^{32}P-labeled 2D12/E2 as described above for genomic Southern blots. After washings and autoradiography, signal intensities of each dilution were quantified on a scanning laser densitometer (Ultrascan XL; Pharmacia LKB).

Chromosome preparations. Five hundred microlitres of heparin-collected whole blood were placed in 10 ml of RPMI 1640 which was supplemented with 18% (v/v) fetal calf serum (GIBCO/BRL) and 0.5% penicillin/streptomycin (GIBCO/BRL). One hundred and fifty microlitres of phytohemagglutinin (M form; GIBCO/BRL) was then added to the blood culture and the culture incubated at 37°C and 5% CO_2 for 72 h. Cell harvesting involved pelleting down all blood cells, resuspending the cells in a 0.075 M KCl hypotonic solution for 20 min at 37°C, followed by 3 fixes in fresh Carnoy fixative (3:1, methanol : glacial acetic acid). Slides were prewashed with 95% (v/v) ethanol, rinsed with distilled water, and soaked in double distilled ice water. Three drops of cell suspension were dropped onto each slide from a height of about 40 cm after which each slide was dried on a hot plate (~70°C) for 5-10 seconds. Slides were aged for at least 5 days in a dessicator at room temperature and then stored at -80°C for up to 1 y before use.

Fluorescence in situ hybridization. Fluorescence *in situ* hybridizations were performed according to Fan *et al.* (1990) with slight modifications. Chromosome preparations were pretreated with DNase-free RNase A in 2x SSC for 1 h at 37°C, and then rinsed twice in 2x SSC at room temperature, dehydrated, and air dried. Chromosomes were then denatured for 3 min in 70% formamide / 2x SSC at 70°C, dehydrated in 2 cold washes of 70% ethanol (at -20°C), and 1 cold wash of 95% ethanol (at -20°C) and air dried. Thereafter, chromosomes were digested with proteinase K (Boehringer Mannheim; 0.06 mg/ml in 2x SSC) for 8 minutes at 37°C, dehydrated, and air dried. One microgram of biotin-16-dUTP labeled 2D12/E2 probe DNA was resuspended in 1 ml of hybridization mixture (50% formamide, 2xSSC, and 100 mg/ml of sonicated salmon sperm DNA), denatured for 5 min at 70°C, applied to each slide, and covered with 24 x 50 mm coverslips. Hybridization proceeded for 16 h in a humid chamber at 37°C. Chromosomes were then washed once in 50% formamide / 2x SSC at 37°C for 10 min and twice in 2xSSC at 37°C for 10 min each.

Subsequent steps for hybridization signal amplification and chromosomal localization were as described in Lin *et al.* (1993). Photomicrographs were taken on FUJI Super G color film, 400 ASA with a Zeiss Axioskop fluorescence microscope.

Results:

Restriction enzyme analysis of the 2D12/E2 clone. Multiple hybridization banding patterns were observed with genomic DNAs separated by standard gel electrophoresis and probed with ^{32}P-labeled 2D12/E2 DNA, revealing the repetitive nature of 2D12/E2 DNA in the human genome (Figure 2.1). The enzymes EcoRI, PstI, and SstI did not produce the typical ladder pattern which is often characteristic of tandemly repetitive DNAs (Horz and Zachau 1977). Furthermore, these three enzymes did not reveal a gain or loss of any particular restriction fragment between 20 kb and < 0.5 kb among the nine individuals examined.

SstI-digested genomic DNAs from 8 unrelated individuals were separated by pulsed-field gel electrophoresis and hybridized to ^{32}P-labelled 2D12/E2, revealing 8 distinct hybridization bands (28 kb, 22 kb, 20 kb, 14 kb, 12.5 kb, 11.5 kb, 9.5 kb, and 6 kb) with the 20 kb bands being the most prominent (Figure 2.2A). SstI-digested genomic DNAs, fractionated under similar conditions but probed with ^{32}P-labelled 50E4 (gamma 8) DNA, produced a hybridization pattern different from that with the 2D12/E2 probe. Two dark hybridization bands of 36.5 kb, and 28 kb and a very faint hybridization band of 32 kb were detected in all 8 DNA samples (Figure 2.2B). HpaI-digested genomic DNAs, separated under comparable conditions and hybridized with 2D12/E2 DNA produced a hybridization pattern consisting of 6 bands: 80 kb, 39 kb, 25 kb, 22 kb, 19 kb, 17 kb, and 12 kb (Figure 2.2C). Likewise, similarly digested and fractionated genomic DNAs probed with ^{32}P-labelled 50E4 (gamma 8) demonstrated a somewhat different hybridization pattern from that seen with the 2D12/E2 probe. Hybridization bands of 39 kb, 22 kb, 19

Figure 2.1. Southern blot hybridization of human genomic DNAs fractionated by conventional gel electrophoresis, and probed with clone 2D12/E2.

EcoRI- and PstI-digested genomic DNAs from three females (individuals *1*, *2*, and *3*) are shown. In addition, SstI-digested genomic DNAs from two of these three females are presented. The marker lane (*M*) contains the DNA from a 1 kb ladder (GIBCO/BRL) and molecular weights of certain DNA fragments are indicated on both sides of the Southern blot.

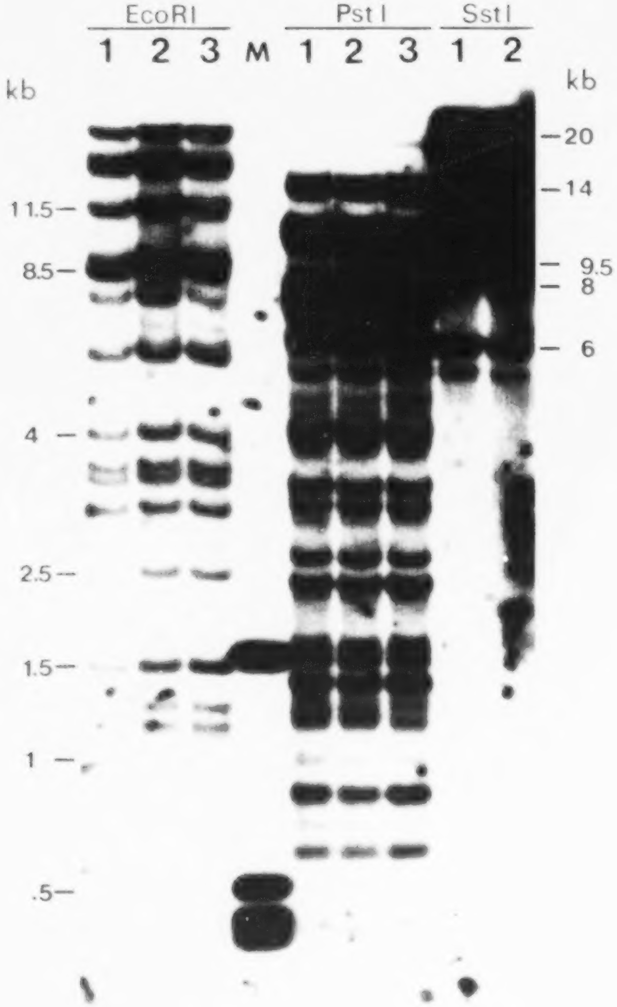
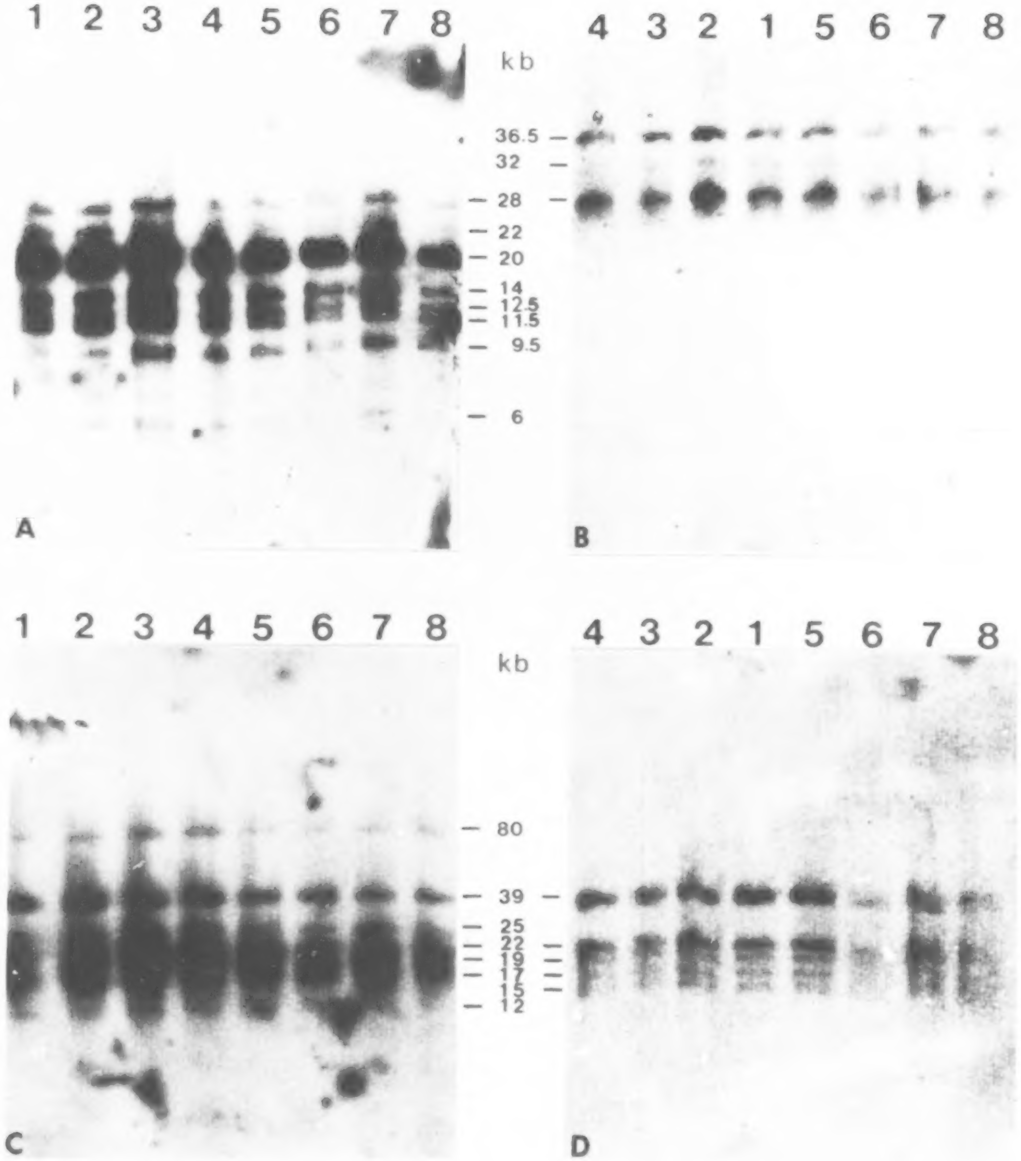

EcoRI
PstI
SstI
1 2 3 M 1 2 3 1 2
kb
11.5
8.5
4
2.5
1.5
1
.5
kb
20
14
9.5
8
6

Figure 2.2. Southern blot hybridizations of human genomic DNAs fractionated by pulsed field gel electrophoresis, and probed with clone 2D12/E2 or 50E4 (gamma-8 satellite DNA).

Individuals 1 and 2 are both unrelated, healthy females and individuals 3-8 are all unrelated healthy males. Genomic DNAs were digested with SstI and probed with (**A**) ^{32}P-labeled 2D12/E2 or (**B**) 50E4. Also, genomic DNAs were digested with HpaI and probed with (**C**) ^{32}P-labelled 2D12/E2 or (**D**) 50E4. Molecular weights of hybridizing DNA fragments are indicated between the Southern blots of similar restriction digestions.


1 2 3 4 5 6 7 8
kb
36.5
32
28
22
20
14
12.5
11.5
9.5
6
A
4 3 2 1 5 6 7 8
B
1 2 3 4 5 6 7 8
kb
80
39
25
22
19
17
15
12
C
4 3 2 1 5 6 7 8
D

kb, 17 kb, and 15 kb were observed, with the 39 kb and 22 kb being most prominent (Figure 2.2D).

DNA sequence analysis of 2D12/E2. The 2D12/E2 subclone was constructed by ligating a 1.2 kb EcoRI fragment from the CX16-2D12 cosmid clone (Fan *et al.* 1992) into pUC 18 vector. Smaller fragments of 2D12/E2 were produced from double digests of EcoRI/SstI and EcoRI/PstI and used as templates for sequencing (Figure 2.3A). Each fragment was sequenced from both of its ends at least three times and the five overlapping DNA sequences were compiled in the construction of the complete 1205 bp DNA sequence of the 2D12/E2 insert (Figure 2.3B).

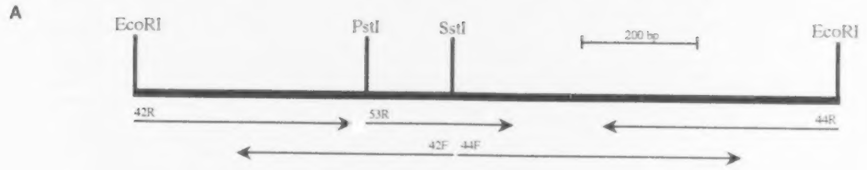
When the complete DNA sequence of 2D12/E2 was analyzed, it was found to contain five essentially complete monomers of 220 bp (b, c, e, f, g) and two partial monomers (a and d) (Figure 2.3B). Monomers a, b, c, and d were arranged in a "tail-to-head" fashion from nucleotides 75, 294, 519, and 585, respectively. Monomers e, f, and g were oriented in a "head-to-tail" fashion, starting from nucleotides 586, 794, and 1015, respectively.

To produce a consensus sequence which is comparable to that published for the gamma 8 satellite DNA monomers, the nucleotide sequences of the opposite strand of the first four monomers (a-d), were aligned to the sequenced strand of the last three monomers (e-g) (Figure 2.4). The DNA sequence similarity of each essentially complete monomer to the derived consensus sequence ranged from 72.8% (monomer g) to 78.6% (monomer b). Most mismatches observed were single base pair deletions, insertions, or substitutions. Exceptions included a 3 bp insertion between nucleotides 179 - 180 of monomer c, a 2 bp deletion of nucleotides 47 - 48 in monomer e, and a 3 bp deletion of nucleotides 74 - 76 in monomer g. No internal duplications were noted within any given monomer.

When the consensus sequence for the 2D12/E2 monomers was aligned to the consensus sequence for the 50E1 and 50E4 monomers, approximately 62% sequence

Figure 2.3. Nucleotide sequence and monomer organization of the 2D12/E2 insert.

(**A**) Restriction map and sequencing strategy for the 2D12/E2 clone. Restriction sites used for subcloning and sequencing are shown. Direction and extent of sequencing are denoted with *arrows*. (**B**) Nucleotide sequence of 2D12/E2. Tandemly organized monomeric repeats of approximately 220 bp are designated *a* to *g* with an arrow showing the beginning and orientation of each monomer. (**C**) Schematic representation of monomer orientations in the 2D12/E2 clone. Two complete and two partial monomers are diagrammed in a tail-to-head orientation from nucleotides 1-585 (*filled in arrows*). Also, three essentially complete monomers are diagrammed in a head-to-tail fashion from nucleotides 586-1205 (*open arrows*). DNA sequencing from the PstI site (*53R*) from the SstI site (*44F*) extends across the point where the 220 bp monomers change orientation.



B

```
   1  GAATTCCGGA TCCAAGTCTA CTCTTCAGCC TTGNTGATTT TTTCCTTGGG NTGCCTCAAC CTCCCCCTGA
                            <-a
  71  AAGTCTGGTG TCCTGTAACT GTTCCCCTAG GGATATCTTC CCACTTTGGG GTGCCCCCCA TGGGAGAAAC
 141  ACACACCCTT GGTGAAAGCC AGGGACTCCA CGACCCCCTT GGGCCCGGCT CTGGAACTGC TGGAAAGGCA
 211  CTGGTCTCTG GTAGAGTACA CCAGCCTCCT ATTCGGACTC GCCACTTTTT TCCTCCAGCT GCCTCAATGT
                                 <-b
 281  CACCCTGAAA TCCTGGCATC GTGACCACGT CCCTCGGGGA CTTCTTGTCA TGTTGGGCGG GCCGCCTGTG
 351  GGGACAGACA CACACACCCT GTGTGAAAGN CCAGGGGACT CCACAACCCC TGCAGGGCCC AGGANAGGAG
 421  CTGCTGGGAA GGTACTGGCC TCCAGTGGAA GACTTCAGNC CTCCATTTCA GCCTCGCCGC TTTTTTCCTC
                                                  <-c
 491  AGGCTGCTTC AACGTACCCC TGAAAGCCTG GNATCCTGTT TCAGTCACTC GGGATACTTA TTNCCACTTT
                            <-de->
 561  GGGGAGCTCC CCCGTAGGAG AGACATGCAC CCTGCATGCG TGGCCTGAGA AAAAAGTGGT GAGGACAAAG
 631  AGGAGACCAG GATGCCCCAC CAGAAGACAG CGCTTTCCCG GCAGCCCCTG CGCATGACCC AGGTGGGTGG
 701  TGGAGTCCCT GACCCCACTA GGGGTGTGTG CCTCTCCCAC AGAAGGCACC ACAAAGCAGC AAGAAGTCCC
         f->
 771  CCAGGGAACA GAGAAAGGCA GCCAGGCTTT CAGAGGGGAG GTTTGAGGCA GGCTTGGGAC AAAATAGGGA
 841  AGGCCAAAAG AGGACGCCTG GGTNCCTTCG CCAGAGGTCA GTGCCTTCCT GGAAGCTCCC GAGCCCGATC
 911  TGGGGGAGTC CTGACGTCCC TGGCCCTCAT GCAGGGTGCG TGTCTCTGCC ACAGAAGGCA CCTTNATGCC
                       g->
 981  GCAAGAAGTC CCCAGGGGAC AGAGACAGGC GGCCAGNCTT TCAGGGGGAG GATGAAGTAG GCCTGGGACA
1051  AAAGCGGCGA GGCCAAACGA GGTCAGGGTG CCCCACAGGA GGTCAGGGCC CTCCTAACAA CCCCTGCGCC
1121  CAAACCGGGG GTGTCGTGGA GTCCCTGGCT CCCACCAGGG GCACGTGTCT CTCCCACCGA GGACACCCCA
1191  AAGCTGCAAG AATTC
```

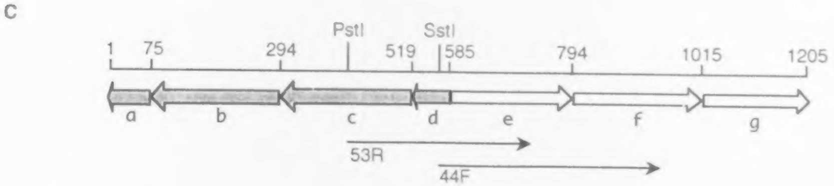


Figure 2.4. Derivation of a 220 bp consensus sequence from the monomers in clone 2D12/E2.

Seven monomeric repeat units are aligned to show maximum DNA sequence similarity. The nucleotide sequences from the opposite strand (*Rev*) of the first four monomers (*Mon/a* - *Mon/d*) and sequenced strand (*Fwd*) of the last three monomers (*Mon/e* - *Mon/g*)the were used in the alignment. The 220 bp consensus sequence is based on the most abundant nucleotide at each position. In the event that at a given position, two different bases were equally prominent, both bases are indicated in the consensus sequence. Nucleotides which are identical to the consensus sequence are replaced with *dots*. Gaps (-) were occasionally introduced to improve the alignment of certain monomers. Ambiguous bases are denoted with an "N" and insertions are indicated by *small vertical arrows* and *lower case letters*.

```
                      10         20         30         40         50         60         70
                       |          |          |          |          |          |          |
                    T                                           T                      A
Consensus     AACAGA ACA GGATGCCAGG CTTTCAGGGG GACGTTGAGG CAGCCN AGG AAAAAAGCGG CGAGGCC AA
                    G                                           G                      G
Mon/a-Rev     .....TT... ...CA....A .......... ..G....... ..N..CA... ......T.AN .A....TG..
Mon/b-Rev     .CGT.GT-.. C......... A........T ...A...... ....TGG... .......T.. ....T..G..
Mon/c-Rev     .C-T..A... ....N..... .......... T.......A. .....TG... .......... ......TG..
Mon/d-Rev                                                                               ->
Mon/e-Fwd                                .G..CCCT .-.A-..C.T GGC.TG--A. .......T.. T....A.A..
                                           ↓a                                         ↓a
Mon/f-Fwd     ......G.A. ..CA...... .......... ..G....... ...G.TTG.. .C....TA.. GA...C.A..
                                                                                      ↓a
Mon/g-Fwd     G.....G... ..CG.....N .......-.. .GA.A...A. T..G.CTG.. .C........ .....C.A.C


                      80         90        100        110        120        130        140
                       |          |          |          |          |          |          |
                             T              C              A            A   C
Consensus     GAGGAGGCCT GGAT CCTCC ACCAGAGG C AGTGCCTTTC C GGCAGCTC CTG GCC GA CCCGGGGGGG
                             G              T              T            C   G
Mon/a-Rev     ...T..A.T. ....-..GGA .TTC
Mon/b-Rev     T.......-. ..TGTA...T .......AC. .......... .A-....T.. .A.A...G.G ...AA.....
                                                                                  ↓t
Mon/c-Rev     AT.....N.. .A.GT.T... ..TG....C. ...A....C. .A-....... ..NT..TG.G ...GCA....
Mon/d-Rev                                                                               ->
Mon/e-Fwd     ......A..A ....G..-.. ......A.A. ..C.-..... .C......C. ...C..AT.. ...A..T...
Mon/f-Fwd     .....C.... ..G.N...T. G.......T. .......-.. .T..A..... .C.A...C.. T.T.....A.
Mon/g-Fwd     ...---GT.A ..G.G.-C.. ..AG....T. ..G...-C.. .TAA..A.C. ...C...CA. A.......T.


                     150        160        170        180        190        200        210
                       |          |          |          |          |          |          |
                                                                A               C
Consensus     TCGTGGAGTC CCTGGCCCTC ACCAAGGGTG TGTGTCTCTC CCACAG GGG CACCCCAAAG  GGCAAGAAG
                                                                G               T
Mon/a-Rev     <-
Mon/b-Rev     .......... ......TT.. .......... .....T.... ...TG.G... .......... T..G....TA
                      ↓c         ↓c                    ↓gtc            ↓g
Mon/c-Rev     .T........ .....N.T.. ..AC...... .....G.... ......GC.. .C......CA T.A.......
Mon/d-Rev     <-
                                                                        ↓c
Mon/e-Fwd     .G........ ....A...-. ..T.G.....   ........ .T..G.G..A GCT....... T..N..T...
Mon/f-Fwd     ..C..AC... .......... .TGC...... ....C..... ......AA.. ....A..... CA........
Mon/g-Fwd     .......... ......T.C. ....G...CA C......... ....C.A..A .......... CT.......T


                     220
                       |
                   C
Consensus     TCCCC AGGG
                   G
Mon/a-Rev     <-
Mon/b-Rev     ....TAG...
Mon/c-Rev     .....G....
              ↓a
Mon/d-Rev     .T...G..T.
Mon/e-Fwd     .....C....
Mon/f-Fwd     .-...C....
Mon/g-Fwd     ..
```

similarity was observed (Figure 2.5). Most mismatches between the two consensus sequences were in the form of base pair substitutions. Single base gaps were required in the gamma 8 consensus sequence (at nucleotide positions 72, 158, 192, 216) and the 2D12/E2 monomer consensus sequence (between nucleotide positions 122-123, 197-198) to optimize DNA sequence similarity. A 20 bp region of 100% sequence similarity was identified between nucleotides 24 - 43 of the two consensus sequences. Based upon a similar repeat size and the sequence similarity observed between the consensus sequences for the 2D12/E2 monomers and the gamma 8 monomers, 2D12/E2 was thought to represent another subfamily of gamma satellite DNA and hence termed gamma X satellite DNA.

Copy number estimation. Slot blot hybridization experiments were used to determine the copy number of gamma X satellite DNA in the human genome. Four micrograms of female genomic DNA produced a similar signal intensity as 0.2 ng of EcoRI-digested 2D12/E2 (data not shown). Using Mandel *et al.*'s (1950) estimation that the human genome contains 3.4 x 10^{-12} g of DNA, it was estimated that gamma X satellite DNA comprises 0.015% of the human genome (i.e. 487, 500 bp in one X-chromosome). This would suggest that the centromere of each X-chromosome harbors approximately 2216 copies of the 220 bp repeat and if contiguous, would span a total array length of approximately 0.5 Mb.

Fluorescence in situ hybridization. Of the twenty five mitotic metaphases examined, FITC fluorescence signals could be observed in the primary constriction of more than 90% of the X-chromosomes. FITC signals were usually seen as two separate fluorescent dots at the lateral sides of the X-chromosome centromere (Figure 2.6). No signal was observed at the primary constriction of the Y-chromosome or any of the autosomes, under the FISH conditions used. The size of the hybridization signal appeared to be much smaller than the

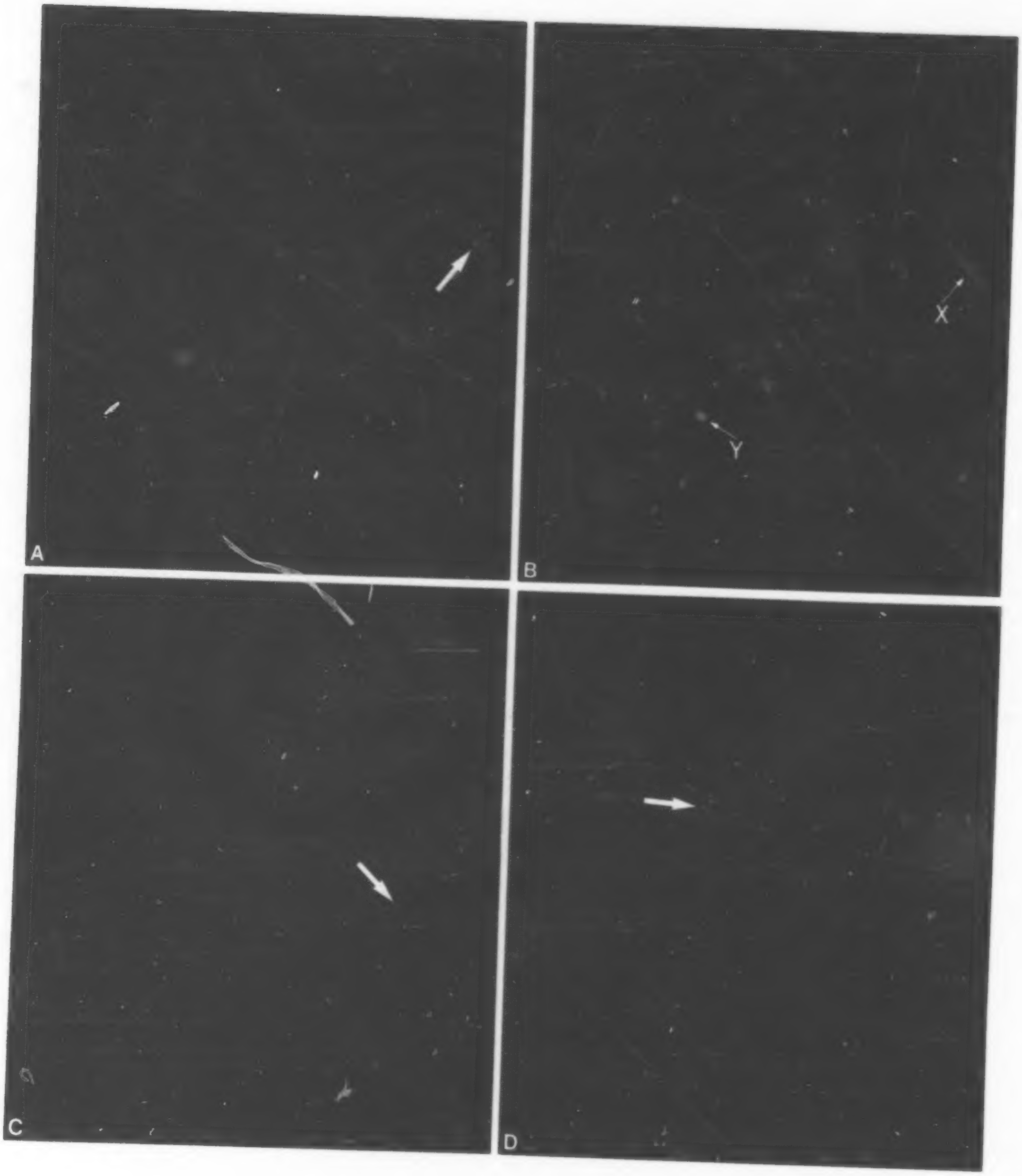
Figure 2.5. Comparison of the 2D12/E2 monomer consensus sequence with the gamma 8 monomer consensus sequence.

The 2D12/E2 monomer consensus sequence (*2D12/E2 con*)was compared to the previously determined gamma 8 DNA consensus sequence (*Gamma 8 con*) (Lin *et al.*, 1993). Gaps (-) were occasionally introduced to improve the alignment and obtain maximum sequence similarity. The nucleotides in the gamma 8 consensus sequence, which were found to be identical to the corresponding nucleotides in the 2D12/E2 monomer consensus sequence, were replaced with *dots*. The largest continuous region of sequence identity can be found between nucleotides 24-43 and is hi-lighted with a box.

```
                        10          20          30          40          50          60          70
                         |           |           |           |           |           |           |
                      G                                               G                        A
2D12/E2 con     AACAGA ACA  GGATGCCAGG  CTTTCAGGGG  GACGTTGAGG  CAGCCN AGG  AAAAAAGCGG  CGAGGCC AA
                      T                                               T                        G
Gamma 8 con     TGGC.TGGGC  ..GCCG....  GAC.......  ..........  ...GCAG...  GG.G......  ....A..GC.

                        80          90         100         110         120          130         140
                         |           |           |           |           |            |           |
                                G               C                A              A   C
2D12/E2 con     GAGGAGGCCT  GGAT CCTCC  ACCAGAGG C  AGTGCCTTTC  C GGCAGCTC  CT-G GCC GA  CCCGGGGGGG
                                T               T                T              C   G
Gamma 8 con     .-...ATG..  ..GAG.....  CAAG....N.  CTCT..CA..  .CA.A.....  .CCAG.A.T.T  .....ACA..

                       150         160           170         180         190          200         210
                         |           |             |           |           |            |           |
                                                                        A
2D12/E2 con     TCGTGGAGTC  CCTGGCCCTC  ACCA--AGGGTG  TGTGTCTCTC  CCACAG GGG  CACCCCA-AAG  NGGCAAGAAG
                                                                        G
Gamma 8 con     CT..AA..C.  ..A....-.T  TGG.GC......  CC...G....  TCG...AA.A  .-.....C...  N.AA..C.G.

                       220
                         |
                     C
2D12/E2 con     TCCCC AGGG
                     G
Gamma 8 con     G..G.-....
```

Figure 2.6. Localization of the 2D12/E2 DNA probe to human male metaphase chromosomes.

(**A**) A metaphase spread from a human male after fluorescence *in situ* hybridization with the biotin labelled 2D12/E2 DNA probe, observed under a filter combination for FITC fluorescence. Hybridization signals (yellow fluorescent dots) can be observed in the primary constriction of a submetacentric C group chromosome (indicated). (**B**) The same metaphase spread observed under a filter combination for 4,6-diamidino-2-phenylindole (DAPI) fluorescence permitting the Q-like banding identification of the afore-mentioned C group chromosome as the X-chromosome (indicated). The Y-chromosome can be readily identified by its size and DAPI fluorescence. (**C**) Another metaphase spread showing FITC hybridization signals after *in situ* hybridization with 2D12/E2 probe DNA. Note the two distinct hybridization signals at the centromere of the X-chromosome (indicated). (**D**) A third metaphase spread again showing FITC hybridization signals of the 2D12/E2 probe to the centromere of the X-chromosome (indicated).


A
B
X
Y
C
D

signals observed with the chromosome X-specific alpha satellite DNA probe, pXBR1 (data not shown). The hybridization signal was comparable in size to that seen with the gamma 8 satellite DNA probe, 50E4 (Lin *et al.* 1993).

Discussion:

Lack of macro restriction fragment length polymorphisms (RFLPs) with HpaI and SstI. The presence of macro RFLPs between unrelated individuals has been well documented in alpha satellite DNA (e.g. Jabs *et al.* 1989; Mahtani and Willard 1990). However, pulsed-field gel electrophoresis studies using essentially 16 different chromosome 8s (probed with gamma 8) and 10 different X-chromosomes (probed with gamma X) failed to reveal any macro RFLPs with either HpaI or SstI. Any intensity differences of a particular hybridization band appeared to be consistent with intensity differences of other hybridization fragments between DNA samples. This suggests that slight differences in the amount of DNA in different agarose plugs could be responsible for the inter-individual differences in hybridization band intensities. This lack of macro RFLPs is restricted to the two different restriction enzymes and eight individuals tested in this study.

Inverted repeats and a conserved 20 bp region. A change in monomer orientation was observed in the 2D12/E2 clone between nucleotides 585 and 586. DNA sequencing from both the PstI site and the SstI site revealed overlapping nucleotide sequences and confirmed the accuracy of the DNA sequence where the repeat units change orientation. Thus the inverted orientation of certain gamma X monomers in this clone appears to be genuine, implying that at least one inversion has likely occurred in the gamma X genomic DNA array. It is believed that this sudden change of monomer orientation at nucleotides 585/586 represents one of the two breakpoints required to produce such an inversion. Identification of the other breakpoint of this inversion could facilitate a size estimation of this inversion.

Inversions in human centromeric satellite DNA have not been frequently documented. Jabs and Persico (1987) first reported a possible inversion in a 3 kb alpha satellite DNA clone from chromosome 6. Wevrick *et al.* (1992) later observed inverted chromosome 7 alpha satellite DNAs in two isolated phage clones. Also, an inversion in the alpha satellite array of the human X centromere was recently identified (Bayne *et al.* 1994). These inversions have all been identified by sequencing DNA clones spanning one of the inversion breakpoints. Genomic inversions of satellite DNA could also be identified by FISH procedures using single stranded oligonucleotides (Meyne *et al.* 1994).

When the gamma 8 and gamma X consensus sequences were aligned with each other, a 20 bp contiguous conserved region was observed. If the 20 bp region is conserved among the different subfamilies of gamma satellite DNA, it is tempting to speculate on its significance. In alpha satellite DNA, a 17 bp conserved motif serves as a binding site for the centromeric protein, CENP-B (Masumoto *et al.* 1989).

Estimations of copy number and array length From slot blot experiments, a 0.015% copy number estimate was obtained for gamma satellite DNA in a single X-chromosome, comparable to the 0.013% estimate previously reported for gamma 8 satellite DNA (Lin *et al.* 1993). Addition of all hybridization fragments observed in pulsed-field fractionated Southern blots suggested an array size of at least 96.5 kb for gamma 8 DNA and 123.5 kb for gamma X DNA. These numbers are believed to be underestimations of the array lengths because of the prominence of certain hybridization bands as a result of the presence of more than one of those fragments per genome (e.g. Figure 2.2A, band 20 kb; Figure 2.2B, bands 22 kb, 19 kb, and 17 kb). The amount of alpha satellite DNA on each human chromosome varies from approximately 800 kb on the Y-chromosome (Cooper *et al.* 1993) to over 3000 kb on the X-chromosome (Jabs *et al.* 1989; Mahtani and Willard 1990).

The amount of gamma satellite DNA found in the centromeric region of either chromosome 8 or X is relatively small compared to that of the other centromeric satellite DNAs. For example, approximately 6 times more alpha satellite DNA (~ 3 Mb) resides in the centromere of the X-chromosome (Jabs *et al.* 1989; Mahtani and Willard 1990). Corneo *et al.* (1967, 1968) estimated that satellite I DNA consisted of approximately 0.5% of the human genome. Assuming a uniform distribution of satellite I DNA amongst the pericentromeric regions of seven different chromosomes, 2 Mb of satellite I DNA is expected in the pericentromeric region of each of these chromosomes. This is four times the amount of gamma satellite DNA found in a single centromere.

Localization and organization of gamma satellite DNA. The discovery of gamma satellite DNA has added another family of satellite DNA to the growing list of human centromeric DNAs. At present, only two subfamilies of gamma satellite DNA have been identified: gamma 8 and X. It has been shown that gamma 8 and X satellite DNA are both restricted to the centromeric regions of their respective human chromosomes, unlike the classical satellite DNAs and beta satellite DNA which are also found in certain non-centromeric heterochromatin and in the short arms of the human acrocentric chromosomes, respectively. Furthermore, unlike alpha satellite DNA, gamma satellite DNA produces small hybridization signals which are often seen as two distinct dots on the lateral sides of the primary constriction, suggesting a close proximity of gamma satellite DNA to the kinetochore domain. Interestingly, Ouspenski and Brinkley (1993) cloned a centromeric DNA sequence from a kinetochore enriched fraction of Chinese hamster ovary cells and found that it hybridized specifically to the centromere of Chinese hamster chromosome 1 as double dots.

The juxtaposed organization of gamma satellite DNA with other satellite DNA families in the centromere should be determined. It has been suggested that alpha satellite DNA is continuous with little or no other intervening DNA sequences (Willard 1990). If this is indeed the scenario, then gamma satellite DNA may be expected to be entirely on one side of or flanking an alphoid DNA array. Further investigations using multiple colored FISH on stretched chromatin (Heng *et al.* 1992; Haaf and Ward 1994) could be useful in determining the juxtaposition of these two centromeric satellite DNA families in human centromeres.

References:

Bayne RAL, Broccoli D, Taggart MH, Thomson EJ, Farr CJ, Cooke HJ. (1994) Sandwiching of a gene within 12 kb of a functional telomere and alpha satellite does not result in silencing. *Hum Mol Genet* 3:539-546.

Brown KE, Barnett MA, Burgtorf C, Shaw P, Buckle VJ, Brown WRA. (1994) Dissecting the centromere of the human Y chromosome with cloned telomeric DNA. *Hum Mol Genet* 3: 1227-1237.

Clarke L. (1990) Centromeres of budding and fission yeasts. *Trends Genet* 6: 150-154.

Cooper KF, Fisher RB, Tyler-Smith C. (1993) Structure of the sequences adjacent to the centromeric alphoid satellite DNA array on the human Y chromosome. *J Mol Biol* 230: 787-799.

Corneo G, Ginelli E, Polli E. (1967) A satellite DNA isolated from human tissues. *J Mol Biol* 23:619-622.

Corneo G, Ginelli E, Polli E. (1968) Isolation of the complementary strands of a human satellite DNA. *J Mol Biol* 33:331-335.

Earnshaw WC, Ratrie H, Stetten G. (1989) Visualization of centromere proteins CENP-B and CENP-C on a stable dicentric chromosome in cytological spreads. *Chromosoma* 101: 322-332.

Fan YS, Davies LM, Shows TB. (1990) Mapping small DNA sequences by fluorescence in situ hybridization directly on banded metaphase chromosomes. *Proc Natl Acad Sci USA* 87: 6223-6227.

Fan YS, Sasi R, Lee C, Court D, Lin CC. (1992) Mapping of 50 cosmid clones isolated from a flow-sorted human X-chromosome library by fluorescence in situ hybridization. *Genomics* 14: 542-545.

Gosden JR, Mitchell AR, Buckland RA, Clayton RP, Evans HJ. (1975) The location of four human satellite DNAs on human chromosomes. *Exp Cell Res* 92: 148-158.

Grady DL, Ratliff RL, Robinson DL, McCanlies EC, Meyne J, Moyzis RK. (1992) Highly conserved repetitive DNA sequences are present at human centromeres. *Proc Natl Acad Sci USA* 89: 1695-1699.

Haaf T, Warburton PE, Willard HF. (1992) Integration of human α-satellite DNA into simian chromosomes: Centromere protein binding and disruption of normal chromosome segregation. *Cell* 70: 681-696.

Haaf T, Ward DC. (1994) High resolution ordering of YAC contigs using extended chromatin and chromosomes. *Hum Mol Genet* 3: 629-633.

Heng HHQ, Squire J, Tsui L-C. (1992) High resolution mapping of mammalian genes by in situ hybridization to free chromatin. *Proc Natl Acad Aci USA* 89: 9509-9513.

Higgins MJ, Wang H, Shtromas I, Haliotis T, Roder JC, Holden JJA, White BN. (1985) Organization of a repetitive human 1.8 kb KpnI sequence localized in the heterochromatin of chromosome 15. *Chromosoma* 93: 77-86.

Horz W, Zachau HG. (1977) Characterization of distinct segments in mouse satellite DNA by restriction nucleases. *Eur J Biochem* 73: 383-392.

Ikeno M, Masumoto H, Okazaki T. (1994) Distribution of CENP-B boxes reflected in CREST centromere antigenic sites on long range α-satellite DNA arrays of human chromosome 21. *Hum Mol Genet* 3:1245-1257.

Jabs EW, Persico HG. (1987) Characterization of human centromeric regions of specific chromosomes by means of alphoid DNA sequences. *Am J Hum Genet* 41: 374-390.

Jabs EW, Goble CA, Cutting GR. (1989) Macromolecular organization of human centromeric regions reveals high-frequency, polymorphic macro DNA repeats. *Proc Natl Acad Sci USA* 86: 202-206.

Johnson DH, Kroisel PM, Klapper HJ, Rosenkranz W. (1992) Microdissection of a human marker chromosome reveals its origin and a new family of centromeric repetitive DNA. *Hum Mol Genet* 1: 741-747.

Kurnit DM, Neve RL, Morton CC, Bruns GAP, Ma NSF, Cox DR, Klinger HP. (1984) Recent evolution of DNA sequence homology in the pericentromeric regions of human acrocentric chromosomes. *Cytogenet Cell Genet* 38: 99-105.

Kurnit DM, Roy S, Stewart GD, Schwedock J, Neve RL, Bruns GAP, Van Keuren ML, Patterson D. (1986) The 724 family of DNA sequences is interspersed about the pericentromeric regions of human acrocentric chromosomes. *Cytogenet Cell Genet* 43: 109-116.

Lin CC, Draper PN, De Braekeleer M. (1985) High resolution chromosomal localization of the β-gene of the human β-globin gene complex by in situ hybridization. *Cytogenet Cell Genet* 39: 269-274.

Lin CC, Sasi R, Lee C, Fan YS, Court D. (1993) Isolation and identification of a novel tandemly repeated DNA sequence in the centromeric region of human chromosome 8. *Chromosoma* 102: 333-339.

Mahtani MM, Willard HF. (1990) Pulsed field gel analysis of α-satellite DNA at the human X chromosome centromere: High-frequency polymorphisms and array size estimate. *Genomics* 7: 607-613.

Mandel P, Metais P, Cuny S. (1950) Les quantites de DNA per leucocyte chez diverses especes de mammiferes. *CR Acad Sci [III]* 231: 1172-1174.

Maniatis T, Fritsch EF, Sambrook J. (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Manuelidis L. (1978) Chromosomal localization of complex and simple repeated human DNAs. *Chromosoma* 66: 23-32.

Masumoto H, Musukata H, Muro Y, Nozaki N, Okazaki T. (1989) A human centromeric antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *J Mol Biol* 109: 1963-1973.

Metzdorf R, Gottert E, Blin N. (1988) A novel centromeric repetitive DNA from human chromosome 22. *Chromosoma* 97: 154-158.

Meyne J, Goodwin EH, Moyzis RK. (1994) Chromosome localization and orientation of the simple sequence repeat of human satellite I DNA. *Chromosoma* 103:99-103.

Mitchell AR, Gosden JR, Miller DA. (1985) p82H: a cloned sequence of the alphoid repeated DNA family found at the centromeres of all human chromosomes. *Chromosoma* 92:369-377.

Mullenbach R, Lutz S, Holzmann K, Dooley S, Blin N. (1992) A non-alphoid repetitive DNA sequence from human chromosome 21. *Hum Genet* 89: 519-523.

Ohashi H, Wakui K, Ogawa K, Okano T, Niikawa N, Fukushima Y. (1994) A stable acentric marker chromosome: possible existence of an intercalary ancient centromere at distal 8p. *Am J Hum Genet* 55:1202-1208.

Ouspenski LL, Brinkley BR. (1993) Centromeric DNA cloned from functional kinetochore fragments in mitotic cells with unreplicated genomes. *J Cell Science* 105: 359-367.

Page SL, Earnshaw WC, Choo KHA, Shaffer LG. (1995) Further evidence that CENP-C is a necessary component of active centromeres: studies of a dic(X;15) with simultaneous immunofluorescence and FISH. *Hum Mol Genet* 4: 289-294.

Vissel B, Nagy A, Choo KHA. (1992) A satellite III sequence shared by human chromosomes 13, 14, and 21 that is contiguous with a satellite DNA. *Cytogenet Cell Genet* 61: 81-86.

Voullaire LE, Slater HR, Petrovic V, Choo KHA. (1993) A functional marker centromere with no detectable alpha-satellite, satellite III, or CENP-B protein: activation of a latent centromere? *Am J Hum Genet* 52: 1153-1163.

Waye JS, Willard HF. (1989) Human b satellite DNA: Genomic organization and sequence definition of a class of highly repetitive tandem DNA. *Proc Natl Acad Sci USA* 86: 6250-6254.

Wevrick R, Willard VP, Willard HF. (1992) Structure of DNA near long tandem arrays of alpha satellite DNA at the centromere of human chromosome 7. *Genomics* 14: 912-923.

Willard HF. (1990) Centromeres of mammalian chromosomes. *Trends Genet* 6: 410-416.

Willard HF. (1992) Centromeres - primary constrictions are primarily complicated. *Hum Mol Genet* 1: 667-668.

CHAPTER 3

ASCERTAINING THE PRESENCE OF GAMMA SATELLITE DNA IN TWO OLD WORLD PRIMATES AND A MITOTICALLY STABLE HUMAN MARKER CHROMOSOME

A version of this chapter is being prepared for publication.

Prologue

Two subfamilies of gamma satellite DNA were identified in the centromeric region of chromosome 8 and X, respectively. *In situ* hybridization studies suggested a close proximity of gamma satellite DNA to the kinetochore domain, implying a structural / functional role for these DNA sequences in the centromere. If indeed so, gamma satellite DNA may be expected to be conserved in the genomes of closely related primate species. Furthermore, if gamma satellite DNA contributes to centromere function, this DNA sequence should be present in mitotically stable marker chromosomes which originate from chromosome 8 or X. Investigations into these possibilities have been conducted and are presented in the following chapter.

Introduction.

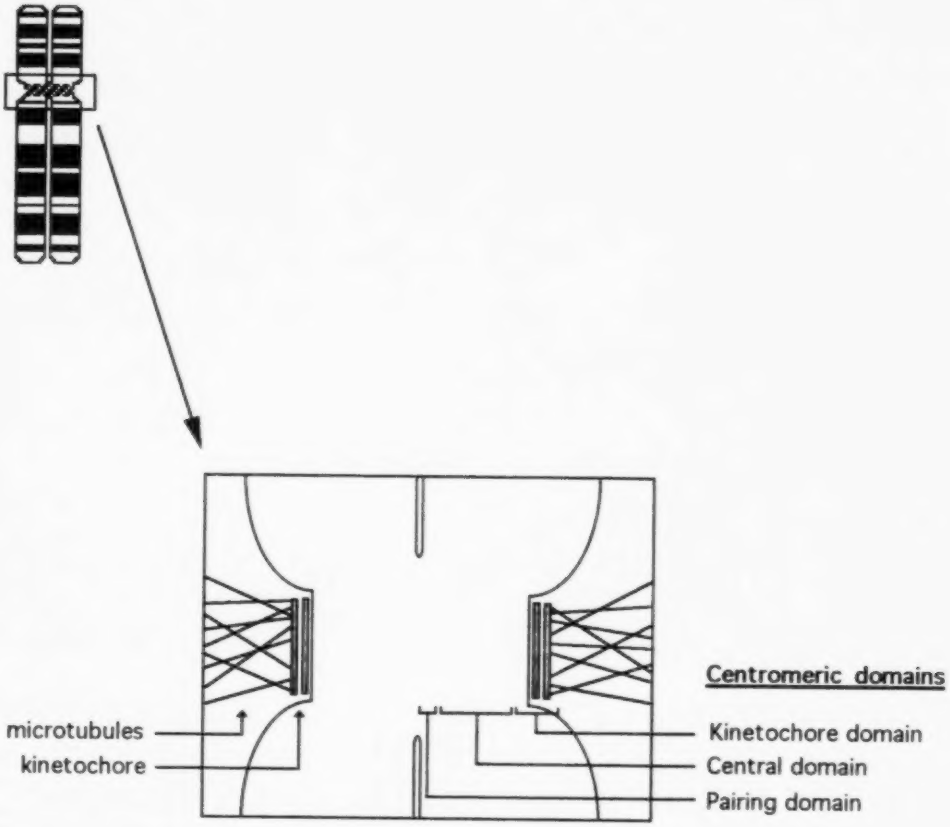
Human centromeric DNAs are primarily composed of different families of tandemly repetitive (satellite) DNA. Among the known human centromeric DNAs, alphoid DNA is by far the best characterized. This DNA family is comprised of 171 bp, AT-rich repeat units. Other tandemly repeated human centromeric DNAs include satellites 1 - 3 (Prosser *et al.* 1986), Sau3A (beta) satellite DNA (Agresti *et al.* 1987, Waye and Willard 1989), and a 48 bp DNA family (Metzdorf *et al.* 1988, Mullenbach *et al.* 1992). More recently, another centromeric DNA sequence was discovered and termed human gamma satellite DNA (Lin *et al.* 1993). This repetitive DNA family was defined by 220 bp, GC-rich tandem repeats, of which two subfamilies were isolated. One subfamily was specific for the centromeric region of chromosome 8 (Lin *et al.* 1993) and the other subfamily was localized to the centromeric region of the X-chromosome (Lee *et al.* 1995).

Human centromeres can be divided into three functional domains: a pairing domain, a central domain, and a kinetochore domain (Figure 3.1). The bulk of centromeric heterochromatin, including alpha satellite DNA and CENP-B proteins (Cooke *et al.* 1990), are confined to the central domain. Fluorescence *in situ* hybridization with gamma satellite DNA probes resulted in two distinct signals at the lateral sides of the primary constriction (Lin *et al.* 1993, Lee *et al.* 1995). This pattern of hybridization signals is usually observed with certain centromeric proteins (e.g. Earnshaw and Rothfield 1985) and satellite DNAs (e.g. Wong and Rattner 1988, Oupenski and Brinkley 1993) thought to be localized to the kinetochore domain. This may imply a close proximity of gamma satellite DNA to the kinetochore and hence a possible involvement in proper kinetochore nucleation. Further characterization of this DNA family is therefore warranted.

Interspecific conservation of repetitive sequences in different primate species may imply functional significance. Alpha satellite DNA has been described throughout the primate order, including the African green monkey (Maio 1971), and new world primates (e.g. Maio *et al.* 1981, Alves *et al.* 1994). The first part of this communication reports on

Figure 3.1. Three functional domains of human centromeres.

The most proximal domain where sister chromatids are in closest contact is referred to as the pairing domain. The central domain is believed to contain the bulk of centromeric heterochromatin, including alpha satellite DNA and CENP-B proteins (Cooke *et al.* 1990). The kinetochore domain is the most distal centromeric region, harboring the trilamellar proteinaceous kinetochore as well as the subjacent chromatin. This figure is adapted from Earnshaw and Rattner (1989).


Centromeric domains
microtubules
kinetochore
Kinetochore domain
Central domain
Pairing domain

investigations of the presence of homologous gamma satellite DNA in the genomes of two old world primates: the African green monkey (*Cercopithecus aethiops*) and the Chimpanzee (*Pan troglodytes*).

Some small yet mitotically stable marker chromosomes, which are formed by chromosome deletions and/or rearrangements, offer another means for characterizing centromeric DNAs. Marker chromosomes which have lost certain centromeric DNA sequences yet retain the ability for proper chromosome segregation imply that the lost sequences are less critical in the formation of a functional centromere (e.g. Voullaire *et al.* 1993, Sacchi *et al.* 1996). Recently, Ohashi *et al.* (1994) also reported a marker chromosome derived from human chromosome 8. Fluorescence *in situ* hybridization studies with a chromosome 8-specific and a non-chromosome-specific alphoid DNA probe failed to detect any alpha satellite DNA in this chromosome. The second part of this communication reports on studies to determine whether gamma 8 satellite DNA is detectable in this mitotically stable marker chromosome.

Materials and Methods.

1. Detection of gamma satellite DNA in two primate species.

Southern blot hybridizations. Genomic DNAs were obtained from the Epstein Barr virus (EBV) -transformed B-cell line of a male chimpanzee (ATCC # CRL1857) and the SV40-transformed COS cell line from the kidney of a male African green monkey (ATCC # 1650) by standard procedures (e.g. Lee and Lin 1996). Six microgram aliquots of genomic DNA from the two primate species were each digested with one of seven different restriction enzymes (i.e. BamHI, EcoRI, HindIII, PstI, RsaI, SstI, or XbaI) and layered into separate wells of 0.8% agarose gels. Electrophoretically fractionated DNAs were transferred to GeneScreen Plus nylon membranes (New England Nuclear) and hybridized to ^{32}P-dCTP-labeled 2D12/E2 (gamma X) DNA (Lee *et al.* 1995). After Southern blot washings and autoradiography, the membranes were stripped of the radioactive probe by washing in

0.1xSSC at 100°C for 20 minutes. The Southern blots were then reprobed with ^{32}P-dCTP-labeled 50E4 insert (gamma 8) DNA (Lin *et al.* 1993), washed, and exposed to autoradiographic film.

Fluorescence in situ hybridizations. Metaphase chromosomes were prepared from the two primate cell cultures as previously described for other mammalian chromosomes (Lee *et al.* 1994). Biotin-16-dUTP (GIBCO/BRL)-labeled 2D12/E2 and 50E4 DNA were used as probes for *in situ* hybridization experiments to African green monkey and chimpanzee chromosomes. *In situ* hybridizations were performed according to a detailed protocol reported previously (Lee *et al.* 1994), and at least twenty separate metaphase spreads were examined for each experiment.

2. Detection of gamma satellite DNA in a mitotically stable marker chromosome.

Fluorescence in situ hybridizations. A lymphoblastic cell line from patient AS (Ohashi *et al.* 1994) was generously provided by Dr. H Ohashi, Division of Medical Genetics, Nagasaki University School of Medicine, Nagasaki, Japan. Biotin-16-dUTP (GIBCO/BRL)-labeled 50E4 and C8-50 DNA were used as probes for *in situ* hybridization studies to metaphase spreads from the lymphoblastic cell line. Immunofluorescent detections were performed as previously described (Lin *et al.* 1993).

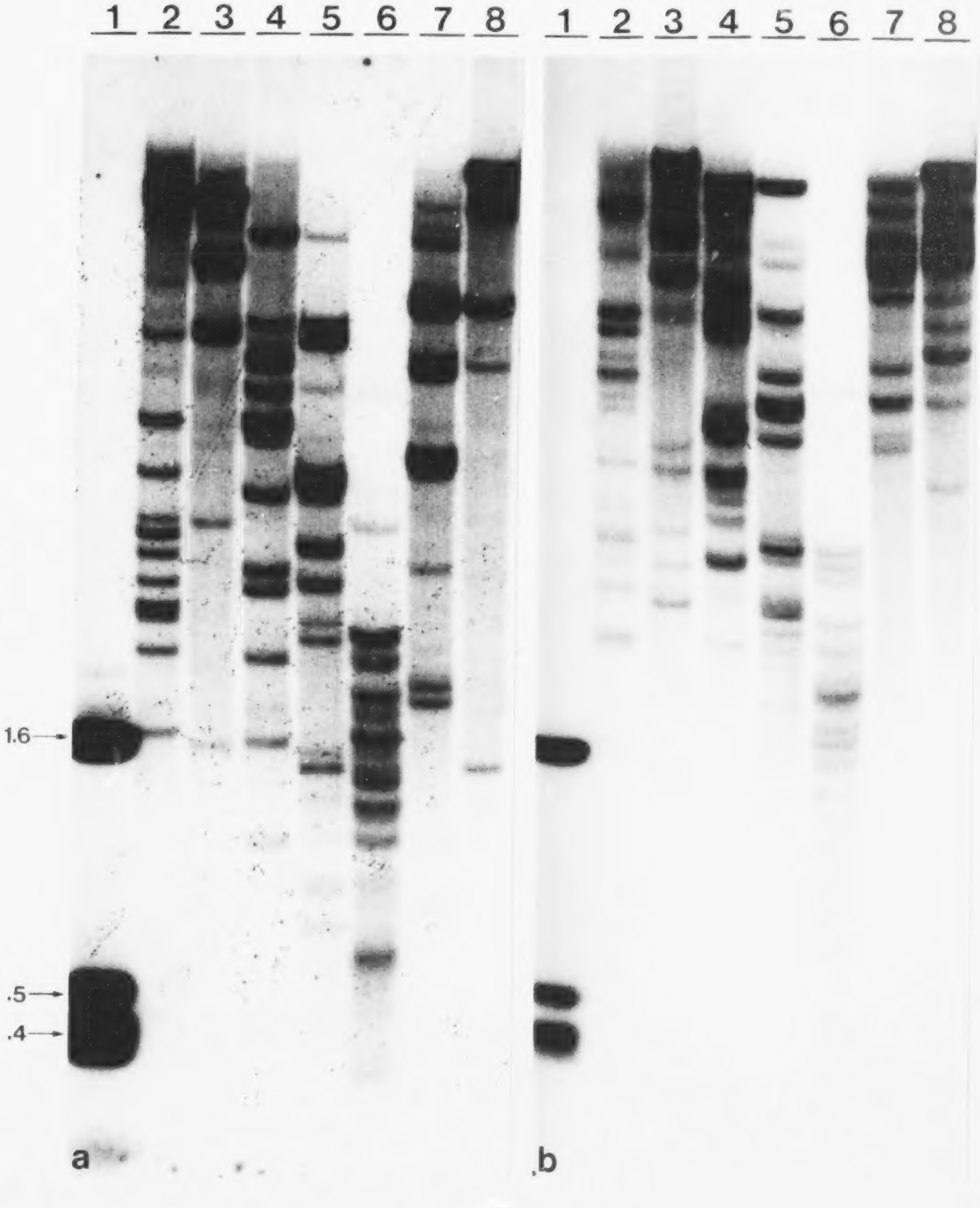
Results

1. Presence of gamma satellite DNA in two primate species.

Southern blot hybridizations. Multiple hybridization bands were observed in each lane of restriction endonuclease-digested African green monkey genomic DNA, when probed with gamma X satellite DNA (Figure 3.2a). Similarily, the same probe produced multiple hybridization bands in each lane of digested chimpanzee genomic DNA (Figure 3.2b). No type A-like ladder pattern was observed with any African green monkey or chimpanzee

Figure 3.2. Southern blot hybridization of gamma X satellite DNA to genomic DNAs from two old world monkeys.

Southern blots containing genomic DNAs from (**a**) a male African green monkey and (**b**) a male chimpanzee were digested with (2) BamHI, (3) EcoRI, (4) HindIII, (5) PstI, (6) RsaI, (7) SstI, and (8) XbaI. Lane 1 of both Southern blots contained DNA fragments from a 1 kb ladder (GIBCO/BRL). Molecular weights of certain marker fragments are indicated (in kb) at the left hand side of panel (**a**) only.


1
2
3
4
5
6
7
8
1
2
3
4
5
6
7
8
1.6
.5
.4
a
b

digested DNAs. Together, this indicated that homologous gamma X satellite DNAs were present as repetitive elements in the genomes of these two primate species. When these Southern blots were stripped of the gamma X probe and hybridized to isotopically-labeled gamma-8 satellite DNA, multiple hybridization banding patterns were again observed for both primate species (data not shown). The banding patterns were different from those seen with the gamma X probe, but still in no distinguishable register.

Fluorescence in situ hybridizations. In the African green monkey chromosomes, human gamma 8 satellite DNA hybridized specifically to the primary constriction of a single pair of submetacentric chromosomes (Figure 3.3a), whereas human gamma X satellite DNA was localized to the centromeric region of a single submetacentric chromosome in diploid cells and two homologous submetacentric chromosomes in tetraploid cells (Figure 3.3b). Hybridization signals in the African green monkey chromosomes appeared consistently larger than the hybridization signals usually seen in human chomosomes.

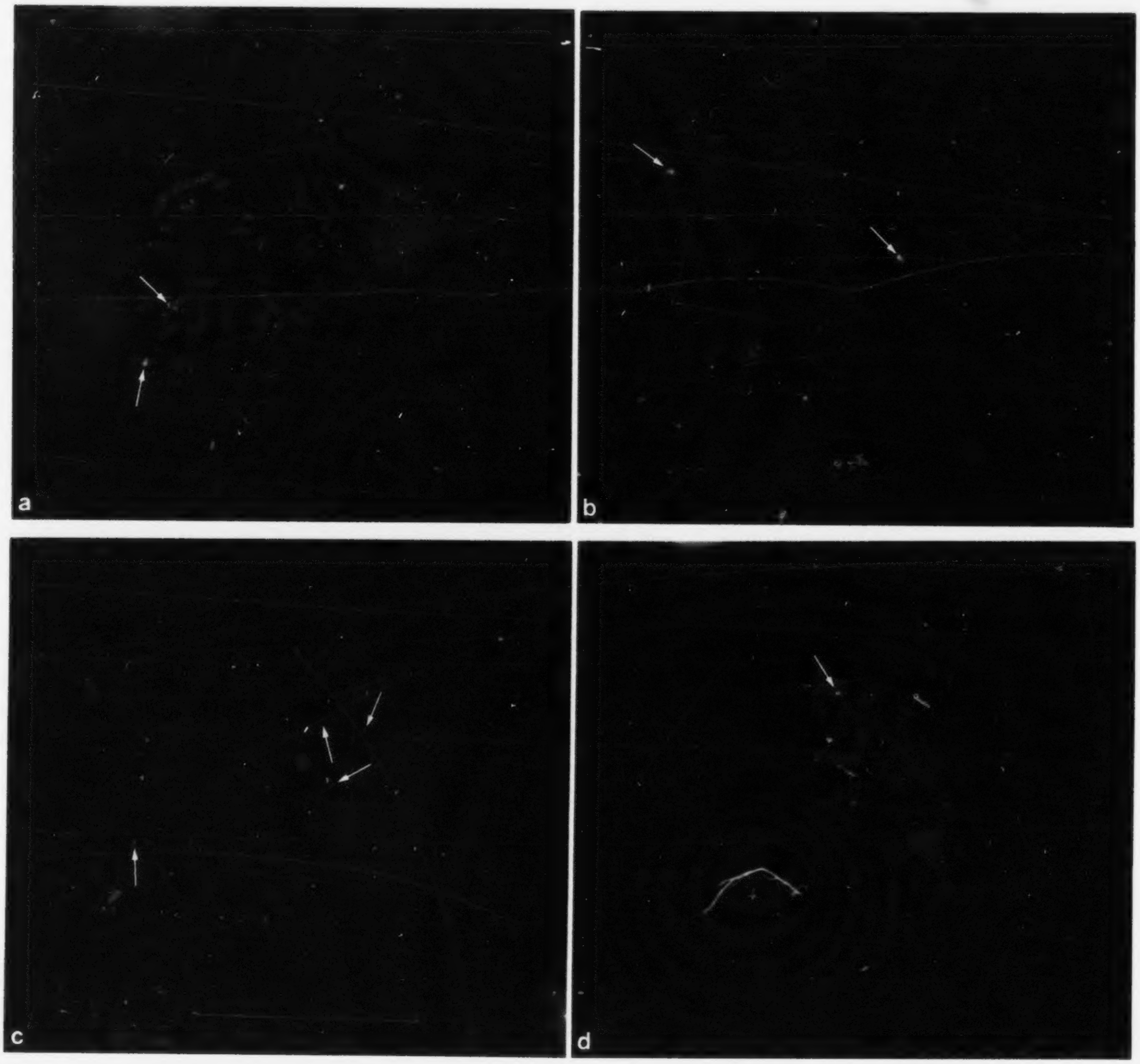
In the chimpanzee chromosomes, predominant hybridization signals with the human gamma 8 satellite DNA probe were confined to the centromeric region of a single pair of submetacentric chromosomes in diploid cells and two pairs of homologous chromosomes in tetraploid cells (Figure 3.3c). Human gamma X satellite probe was localized to the centromeric region of a single submetacentric chimpanzee chromosome (Figure 3.3d). Hybridization signal size and intensity approximated those seen in human chromosomes, and were often seen as two distinct opposing fluorescent dots.

2. Apparant absence of gamma satellite DNA in a mitotically stable marker chromosome.

Fluorescence in situ hybridizations. Examination of forty metaphase spreads from the lymphoblastic cell line showed chromosome complements consistent with the reported karyotype of 47, XX, +mar for patient A.S (Ohashi *et al.* 1994). In each metaphase spread, a single marker chromosome was observed as the smallest chromosome in the

Figure 3.3. Localization of human gamma 8 and X satellite DNAs to metaphase chromosomes of a male African green monkey and a male chimpanzee.

Human gamma 8 satellite DNA was localized to the primary constriction of (**a**) a single pair of submetacentric chromosomes in a diploid African green monkey cell and (**c**) two pairs of submetacentric chromosomes in a tetraploid chimpanzee cell. Human gamma X satellite DNA probe is localized to the centromeric region of (**b**) two homologous submetacentric chromosomes in a tetraploid male African green monkey cell and (**d**) a single submetacentric chromosome in a diploid male chimpanzee cell.


a
b
c
d

complement. Biotin-labeled gamma 8 satellite DNA probe consistently revealed hybridization signals at the centromeric region of a pair of chromosome 8s. No hybridization signal was observed in any other chromosome, including the marker chromosome (Figure 3.4).

Discussion:

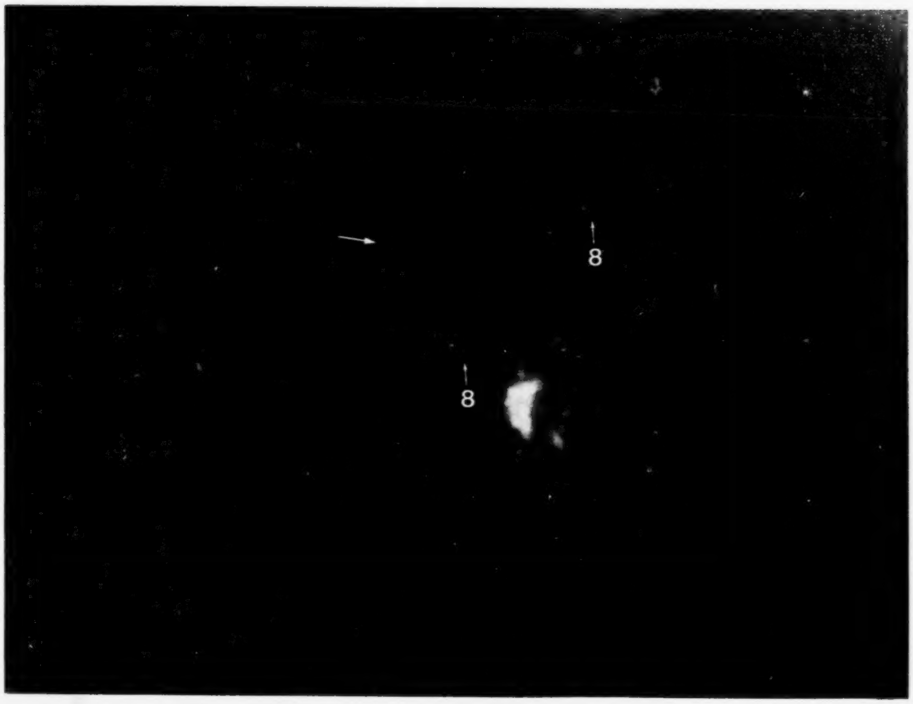
Gamma satellite DNAs are present in the genomes of the African green monkey and chimpanzee. Most primate species fall into three suborders: (1) Catarrhini (old world primates), (2) Platyrrhini (new world primates), and (3) Strepsirhini (prosimians). Humans, chimpanzee, and African green monkeys are all considered members of the suborder Catarrhini and have been shown in the present study to contain gamma satellite DNA sequences. Multiple hybridization bands in Southern blots demonstrated the repetitive nature of gamma satellite DNA in the chimpanzee and the African green monkey. However, type-A like hybridization patterns were not observed with any of the seven restriction enzymes used. Therefore, a 220 bp repeat unit size could not be confirmed for this DNA family in these two primate species.

Fluorescence *in situ* hybridization studies localized the human gamma 8 sequences to the centromeric region of a single pair of submetacentric autosomes in the African green monkey and in the chimpanzee. It is possible that these chromosomes are homologous to the human chromosome 8. The human gamma X probe was localized to the centromeric region of a single submetacentric chromosome in diploid cells of the two primate species. As both cell lines were derived from somatic cells of male animals, the chromosomes in question are likely X-chromosomes.

Interestingly, fluorescent signals were often observed as two distinct dots at the lateral sides of the primary constriction. Resolution of the two hybridization signals at a single centromere was sometimes more difficult in the African green monkey chromosomes. This could be attributed to the overlapping of larger signals seen in this

Figure 3.4. Localization of gamma 8 satellite DNA to metaphase chromosomes from a lymphoblast cell line containing a single mitotically stable chromosome 8-derived marker chromosome.

Fluorescent hybridization signals were observed in the centromeric region of both chromosome 8s (indicated) but not in the small marker chromosome (large arrow).


8
8

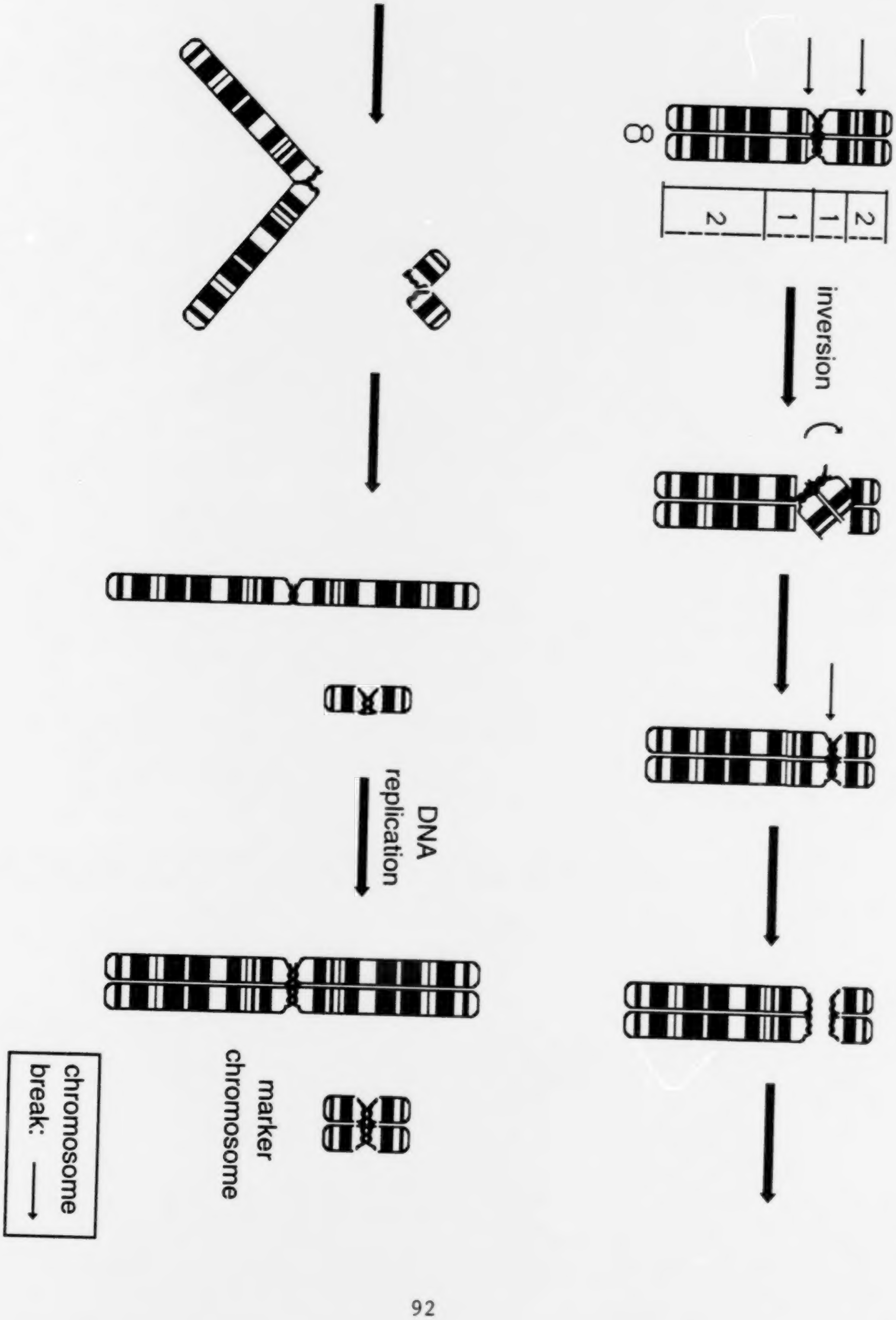
species' chromosomes, possibly as a result of increased array sizes as compared to those in humans. It has already been shown that alpha satellite DNA constitutes a greater portion of the African green monkey genome than the human genome (Maio 1971, Vissel and Choo 1987).

Gamma 8 satellite DNA is undetectable in a mitotically stable human marker chromosome. Mitotically stable marker chromosomes offer an excellent means for determining which centromeric DNA sequences are essential in the formation of a functional centromere. The marker chromosome described by Ohashi *et al.* (1994) had three interesting features: (1) Reverse chromosome painting and two-color FISH studies identified the origin of the marker chromosome as 8pter -> 8p23.1. Amplified DNA from the marker chromosome did not result in any hybridization signal at any centromeric region, including that of chromosome 8. (2) Antikinetochore-antibody immunostaining, using the serum of a patient with the CREST variety of scleroderma, implied the presence of one or more kinetochore protein(s) in this marker chromosome. (3) Alpha satellite DNA could not be detected in the marker chromosome during FISH studies with a chromosome 8-specific and a non-chromosome-specific alpha satellite DNA probe.

Ohashi *et al.* (1994) suggested that this marker chromosome was produced from an inverted duplication of chromosome region 8pter -> 8p23.1 and that a latent centromere in this chromosome region may have become activated during the development of this marker chromosome, conferring mitotic stability. Alternatively, a pericentric inversion, accompanied by a chromosome break in the centromeric region, could also have resulted in the formation of this marker chromosome (Figure 3.5). If this marker chromosome was produced by such a chromosomal rearrangement, then some chromosome 8 centromeric DNAs would be expected in the marker chromosome. The inability to detect chromosome 8 alpha and gamma satellite DNAs by FISH may suggest that either this chromosome is not formed in the manner proposed in Figure 3.5 or that other DNA sequences may contribute

Figure 3.5. A proposed mechanism for the formation of the marker chromosome from patient AS.

Two breaks occur in chromosome 8 which allow a pericentric inversion to occur. This is followed by another break in the centromeric region of the rearranged chromosome 8 resulting in two isochromosomes. The marker chromosome is the smaller of the two isochromosomes and is subsequently inherited by patient AS.


8
2
1
1
2
inversion
DNA replication
marker chromosome
chromosome break: →

to a functional centromere. These studies do not exclude the possibility that alpha and gamma satellite DNAs may still exist in this marker chromosome, but in amounts below the detection sensitivity of FISH.

References:

Agresti A, Rainaldi G, Lobbiani A, Magnani I, Lernia RD, Meneveri R, Siccardi AG, Ginelli E. (1987) Chromosomal location by in situ hybridization of the human Sau3A family of DNA repeats. *Hum Genet* 75: 326-332.

Alves G, Seuánez HN, Fanning T. (1994) Alpha satellite DNA in neotropical primates (Platyrrhini). *Chromosoma* 103: 262-267.

Cooke CA, Bernat RL, Earnshaw WC. (1990) CENP-B: A major human centromere protein located beneath the kinetochore. *J Cell Biol* 110: 1475-1488.

Earnshaw WC, Rothfield NF. (1985) Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma* 91: 313-321.

Earnshaw WC, Rattner JB. (1989) *A map of the centromere (primary constriction) in vertebrate chromosomes at metaphase*. In: Resnick MA, Vig BK, eds. Progress in Clinical and Biological Research, Mechanisms of chromosome distribution and aneuploidy. Reno, Nevada. pp. 33-42.

Lee C, Lin CC. (1996) Conservation of a 31 bp bovine subrepeat in centromeric satellite DNA monomers of *Cervus elaphus* and other cervid species. *Chromosome Res* 4: 428-436.

Lee C, Ritchie DBC, Lin CC. (1994) A tandemly repetitive, centromeric DNA sequence from the Canadian woodland caribou (*Rangifer tarandus caribou*): its conservation and evolution in several deer species. *Chromosome Res* 2: 293-306.

Lee C, Li X, Jabs EW, Court DR, Lin CC. (1995) Human gamma X satellite DNA: An X chromosome specific centromeric DNA sequence. *Chromosoma* 104: 103-112.

Lin CC, Sasi R, Lee C, Fan YS, Court D. (1993) Isolation and identification of a novel tandemly repeated DNA sequence in the centromeric region of human chromosome 8. *Chromosoma* 102: 333-339.

Maio JJ. (1971) DNA strand reassociation and polyribonucleotide binding in the African Green monkey, *Cercopithecus aethiops*. *J Mol Biol* 56: 579-595.

Maio JJ, Brown FL, Musich PR. (1981) Toward a molecular paleontology of primate genomes. 1. The HindIII and EcoRI dimer families of alphoid DNAs. *Chromosoma* 83: 103-125.

Metzdorf R, Göttert E, Blin N. (1988) A novel centromeric repetitive DNA from human chromosome 22. *Chromosoma* 97: 154-158.

Mullenbach R, Lutz S, Holzmann K, Dooley S, Blin N. (1992) A non-alphoid repetitive DNA sequence from human chromosome 21. *Hum Genet* 89: 519-523.

Ohashi H, Wakui K, Ogawa K, Okano T, Niikawa N, Fukushima Y. (1994) A stable acentric marker chromosome: Possible existence of an intercalary ancient centromere at distal 8p. *Am J Hum Genet* 55: 1202-1208.

Oupenski LL, Brinkley BR. (1993) Centromeric DNA cloned from functional kinetochore fragments in mitotic cells with unreplicated genomes. *J Cell Science* 105: 359-367.

Prosser J, Frommer M, Paul C, Vincent PC. (1986) Sequence relationships of three human satellite DNAs. *J Mol Biol* 187: 145-155.

Sacchi N, Magnani I, Fuhrman-Conti AM, Monard SP, Darfler M. (1996) A stable marker chromosome with a cryptic centromere: evidence for centromeric sequences associated with an inverted duplication. *Cytogenet Cell Genet* 73: 123-129.

Vissel B, Choo KH. (1987) Human alpha satellite DNA - consensus sequence and conserved regions. *Nucleic Acids Res* 15: 6751-6752.

Voullaire LE, Slater HR, Petrovic V, Choo KHA. (1993) A functional marker centromere with no detectable alpha-satellite, satellite III, or CENP-B protein: activation of a latent centromere? *Am J Hum Genet* 52: 1153-1163.

Waye JS, Willard HF. (1989) Human β satellite DNA: Genomic organization and sequence definition of a class of highly repetitive tandem DNA. *Proc Natl Acad Sci USA* 86: 6250-6254.

Wong AKC, Rattner JB. (1988) Sequence organization and cytological localization of the minor satellite of mouse. *Nucleic Acids Res* 16: 11645-11661.

CHAPTER 4

CONSERVATION OF A 31 BP BOVINE SUBREPEAT IN CERVID CENTROMERIC SATELLITE DNA

A version of this chapter has been published:

Lee C, Lin CC. (1996) Conservation of a 31 bp bovine subrepeat in centromeric satellite DNA monomers of *Cervus elaphus* and other cervid species. *Chromosome Research* 4: 428-436.

Prologue

In contrast to human centromeric DNAs, very little is known of centromeric DNA sequences of other mammalian species. In deer species (Family Cervidae), karyotypic evolution has resulted in drastic difference in chromosome numbers: from 2n=80 in the Siberian roe deer to 2n=6/7 in the Indian muntjac. Such karyotypic changes provide excellent systems for studying mammalian centromeric structure and function. To date, only one centromeric satellite DNA has been identified in deer species. This DNA family is organized as 0.8 kb repeat units in plesiometacarpalial deer and 1 kb repeat units in telemetacarpalial deer. Two reports on 0.8 kb centromeric satellite DNA monomers from the Indian muntjac suggested the presence of 31 bp subrepeats. Subsequent reports on monomer clones from the Chinese muntjac, European roe deer, and Canadian woodland caribou have all failed to detect such an internal periodicity. Further characterization of this cervid centromeric satellite DNA family is needed to resolve this discrepancy. The following chapter describes the identification of 31 bp subrepeats in all previously isolated cervid centromeric satellite DNA monomers as well as in a new centromeric satellite DNA clone from the European red deer.

Introduction.

It has been estimated that over 95% of some mammalian genomes consist of DNA sequences which do not code for proteins (Nowak 1994). A substantial fraction of these non-coding DNA sequences are repetitive in nature and are either tandemly organized (as in satellite DNAs) or interspersed among other genomic DNA sequences. The repeat unit (monomer) size of certain satellite DNA families can often be determined after specific restriction endonuclease digestions which convert a tandemly repetitive DNA array into a ladder pattern of fragments having sizes equal to integral multiples of the basic repeat unit length.

Extensive DNA sequence analyses of monomers can sometimes reveal the existence of smaller internal repeat units (subrepeats). For example, the 234 bp monomer of mouse major satellite DNA was shown to consist of four internal 58 bp subrepeats. These subrepeats could be further divided into two related 28 and 30 bp segments, each thought to originate from mutations and amplifications of three similar 9 bp nucleotide sequences (Horz and Altenburger 1981).

As much as eight different families of satellite DNA sequences have been identified in the bovine genome based on differential buoyant densities in CsCl equilibrium gradients (reviewed in Singer 1982). Many of these repetitive DNAs (including satellites 1.720, 1.706, and 1.711a) contain subrepeats of 23 bp, which are thought to have resulted from a duplication of an original 12 bp DNA sequence (Pech *et al*. 1979). Bovine satellite 1.715 (also referred to as bovine satellite I), has 1400 bp monomers (Gaillard *et* al. 1981) which consist of tandemly reiterated 31 bp subrepeats (Plucienniczak *et al.* 1982). These 31 bp subrepeats do not appear to share substantial sequence similarity to the 23 bp subrepeat found in the other bovine satellite DNAs although it has been suggested that both subrepeats may have arisen from a common ancestral sequence (Taparowsky and Gerbi 1982, Jobse *et al*. 1995).

In cervid species, a major centromeric satellite DNA family appears to be predominantly organized into monomer sizes of 0.8 kb in the Indian and Chinese muntjac genomes (Bogenberger *et al*. 1985, Yu *et al*. 1986, Lin *et al*. 1991), 1 kb in the reindeer (Lima-de-Faria *et al*. 1984, Lee *et al*. 1994) and 2 kb in the European roe deer (Scherthan 1991). The first two cervid centromeric satellite DNA monomers to be cloned (both 0.8 kb in length and from the Indian muntjac genome) demonstrated the presence of internal 31 bp, tandemly organized subrepeats which were homologous to the 31 bp subrepeat found in bovine 1.715 satellite DNA. Subsequent cloning reports of centromeric satellite DNA monomers from other deer species (i.e. the Chinese muntjac, the European roe deer, and the Canadian woodland caribou) all failed to demonstrate the presence of any internal subrepeats. In the present study, an attempt was made to resolve this discrepancy by cloning and characterizing a centromeric satellite DNA monomer from another deer species, the European red deer. Molecular characterization of this DNA clone (Ce-Pst1) included a DNA sequence analysis strategy for examining the existence of internal direct subrepeats. Internal 31 bp subrepeats were identified in the Ce-Pst1 clone and subsequently in the C5, CCSatI, and Rt-Pst3 clones of the Chinese muntjac, European roe deer, and Canadian woodland caribou, respectively. Thus, amplification of an original 31 bp DNA sequence could have contributed to the genesis of centromeric satellite DNA monomers in cervid species. Moreover, these findings substantiate the notion that amplification of the 31 bp DNA sequence occurred in an ancestral species common to both cervids and bovids.

Materials and Methods.

Cloning of a centromeric satellite DNA monomer from the European red deer.

A cell line derived from testis cells of a male European red deer (*Cervus elaphus hippelaphus*) was kindly provided by Drs. F. Fontana and M. Rubini, University of Ferrara, Italy, and grown in DMEM media (GIBCO/BRL) supplemented with 10% fetal calf serum and 0.5% penicillin/streptomycin. Cells were harvested with 1x trypsin / EDTA (GIBCO/BRL), washed once in PBS, and lysed with a solution containing 100 mM Tris-HCl (pH=8.0), 40 mM EDTA, 0.5% SDS, to extrude genomic DNA. After phenol and chloroform extractions, genomic DNA was ethanol precipitated and redissolved in an appropriate volume of low TE buffer. Twenty micrograms of male red deer genomic DNA were then digested with PstI and fractionated through a 0.8 % agarose gel. A bright ethidium bromide staining DNA band of 0.8 kb was observed and electroeluted, from which a 0.8 kb fragment was cloned into pUC19 as previously described (Lin *et al.* 1991). This clone was designated as Ce-Pst1 (*Cervus elaphus hippelaphus* - PstI clone 1) and further characterized.

Characterization of the Ce-Pst1 clone.

1. Southern blot analyses. Ten microgram aliquots of red deer genomic DNA were each digested with one of 5 different restriction endonucleases: BamHI, HpaII, MspI, PstI, and RsaI. Aliquots were layered into separate wells of a 0.8% agarose gel and electrophoretically fractionated. Southern blot hybridization, filter washings, and autoradiography were performed as previously described (Lee *et al.* 1994).

2. Fluorescence in situ hybridization. Metaphase spreads of male red deer chromosomes were obtained from the above mentioned cell culture following conventional acetic acid / methanol fixation methods. Chromosome preparations were pretreated with RNase A (100 µg/ml), denatured in 70% formamide / 2x SSC for 3 min at 70°C, and digested with proteinase K (0.06 µg/ml). Fluor-12-dUTP (Stratagene) labeled Ce-Pst1 probe DNA was suspended in a mixture of 50% formamide, 2x SSC, and salmon sperm DNA (50 µg/ml), at a concentration of 250 ng/ml. Following denaturation, the DNA probe mixture was applied to the chromosome spreads. Hybridization proceeded overnight in a humid chamber at 37°C, followed by a single 10 min wash in 50% formamide / 2x SSC at 40°C and two 10 min washes in 2x SSC at 40°C. Slides were then immediately mounted in glycerol containing P-phenylenediamine dihydrochloride (1 mg/ml), 4'-6-diamidino-2-phenylindole (DAPI; 0.8 μg/ml) and propidium iodide (PI; 0.4 μg/ml). Chromosomes were viewed on a Zeiss Axioskop fluorescent microscope and photographed with Kodak Gold Plus, ASA 400 film.

3. DNA sequencing. The Ce-Pst1 clone was sequenced from both ends using the dideoxy chain termination Sequenase kit (United States Biochemical Corporation) and each sequencing reaction was repeated twice. The DNA sequence data was compiled on a DNA Sequence Editor Software Program (Applied BioSystems) and the complete DNA sequence of the Ce-Pst1 clone was deposited into the GenBank database (Accession number : U48429).

Source of other cervid centromeric satellite DNA monomer sequences.

The other cervid centromeric satellite DNA sequences used for sequence comparisons and internal repeat analyses were all previously reported: the 1A clone of the Indian muntjac (*Muntiacus muntjak vaginalis*) (Bogenberger *et al.* 1985; EMBL accession number X02323), the C5 clone of the Chinese muntjac (*Muntiacus reevesi*) (Lin *et al.* 1991; EMBL accession number X56823), the CCSatI clone of the European roe deer (*Capreolus capreolus capreolus*) (Scherthan 1991), and the Rt-Pst3 clone of the Canadian woodland caribou (*Rangifer tarandus caribou*) (Lee *et al.* 1994; EMBL accession number X77013).

Internal repeat analyses.

Internal repeat analyses were performed on each available cervid centromeric satellite DNA monomer sequence in a manner similar to that described by Plucienniczak *et al.* (1982). Specifically, each cervid monomer was compared to a DNA sequence consisting of two adjacent copies of the same monomer. The single copy cervid monomer was then shifted to the right in 1 base increments, with respect to the two-copy DNA sequence. The total number of identical nucleotides observed between the two aligned DNA sequences, after each sequence shift, was plotted on a line graph using the CA-Cricket Graph III program (Computer Associates). If a monomer contained an internal 31 bp periodicity, the internal subrepeats in both DNA sequences should align and become "in frame" with each other after every 31 base shifts, producing a significant increase in the number of identical nucleotides and resulting in a peak in the line graph. An Average Number of Identical Nucleotides (ANIN) was then calculated for each monomer by dividing the sum of all data from nucleotide shift self-comparisons by the total number of shifts for the monomer, excluding values for "in-frame" peaks. The percent increase of each "in-frame" peak from the ANIN was calculated by using the following formula: ((Peak size - ANIN) / ANIN) x 100%.

If a 31 bp periodicity could be detected in these satellite DNA monomers, the bovine 1.715 satellite subrepeat consensus sequence could be used to arbitrarily define the boundaries of the individual 31 bp subrepeats within each cervid monomer (as described by Bogenberger *et al*. 1985). Subtraction of certain nucleotides and insertion of gaps were occasionally required to improve the alignment of the internal subrepeats. A 31 bp consensus sequence could then be constructed for each cervid monomer based on the most frequently occurring nucleotide at each position of the aligned subrepeats.

Sequence comparisons were also made between each monomer-specific 31 bp cervid consensus sequence and the bovine 31 bp subrepeat consensus sequence. As well, each monomer-specific 31 bp cervid consensus sequence was examined for the presence of any further internal repeats.

Results.

Characterization of the European red deer centromeric satellite DNA clone, Ce-Pst1:

Hybridization of the Ce-Pst1 DNA probe to a Southern blot of red deer genomic DNAs revealed multiple hybridization bands with prominent bands consistently arranged in a 0.8 kb register for all five restriction enzyme digestions (Figure 4.1). This suggested that the majority of Ce-Pst1 DNA is organized as tandemly arranged 0.8 kb repeat units in the red deer genome. Light hybridization bands, in a 0.7 kb and a 0.9 kb register, could also be seen with the BamHI-digested DNA (Figure 4.1, lane 1). More prominent hybridization bands in similar 0.7 kb and 0.9 kb registers were observed in the HpaII and MspI digested DNAs (Figure 4.1, lanes 2 and 3), implying that a substantial amount of Ce-Pst1 DNA could also be organized into monomers of 0.7 kb and 0.9 kb. Other faint bands could also be observed in the PstI- and RsaI-digested DNAs (Figure 4.1, lanes 4 and 5), but in no distinguishable register.

Examination of 30 metaphase chromosome spreads from male red deer cells revealed a karyotype of 2n=68 consisting of 64 acrocentric autosomes, a pair of metacentric autosomes, a large acrocentric X-chromosome (the largest acrocentric chromosome of the complement) and a small submetacentric Y-chromosome (Figure 4.2). These observations were consistent with the male red deer karyotype reported by Gustavsson and Sundt (1968). Hybridization of the fluorescein-labeled Ce-Pst1 probe to the metaphase chromosomes revealed large bright fluorescent signals at the centromeric region of all acrocentric chromosomes, including the X-chromosome (Figure 4.2A). The single pair of metacentric autosomes and the submetacentric Y-chromosome did not appear to exhibit any hybridization signal.

DNA sequencing of the Ce-Pst1 clone revealed a nucleotide sequence of 806 bp (Figure 4.3A) which was slightly GC-rich (54.2%). Remarkable features of the DNA sequence included a contiguous run of 8 cytosines at nucleotides 403 - 410, and the occurrence of 15 CAGG or GAGG tetranucleotides. When the Ce-Pst1 sequence was

Figure 4.1. Tandemly repetitive organization of the Ce-Pst1 DNA sequence in the genome of the European red deer (*Cervus elaphus hippelaphus*).

A Southern blot of red deer genomic DNA digested with BamHI (1), HpaII (2), MspI (3), PstI (4), RsaI (5), and probed with ^{32}P-labeled Ce-Pst1 insert DNA. DNA fragment sizes of 0.7 kb, 0.8 kb, and 0.9 kb are indicated to the left of the figure.

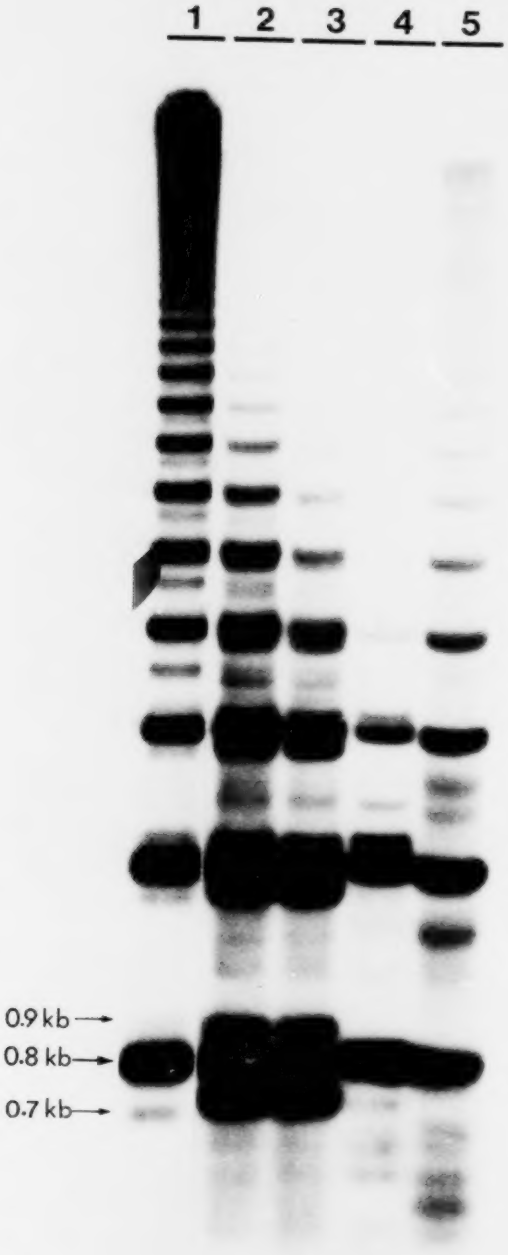

1
2
3
4
5
0.9 kb
0.8 kb
0.7 kb

Figure 4.2. Localization of the Ce-Pst1 satellite DNA to the chromosomes of a male European red deer by fluorescence *in situ* hybridization.

(**A**) A metaphase spread of a male European red deer observed with a filter combination for FITC fluorescence after hybridization and immunofluorescent detection of biotinylated labeled Ce-Pst1 DNA probe. The centromeric region of all acrocentric chromosomes, including the X-chromosome (indicated) showed large bright hybridization signals. A single pair of metacentric autosomes (denoted by small arrows) and the Y-chromosome (indicated) did not display any hybridization signal. (**B**) The same metaphase spread (as in *A*) observed with a filter combination for DAPI fluorescence.

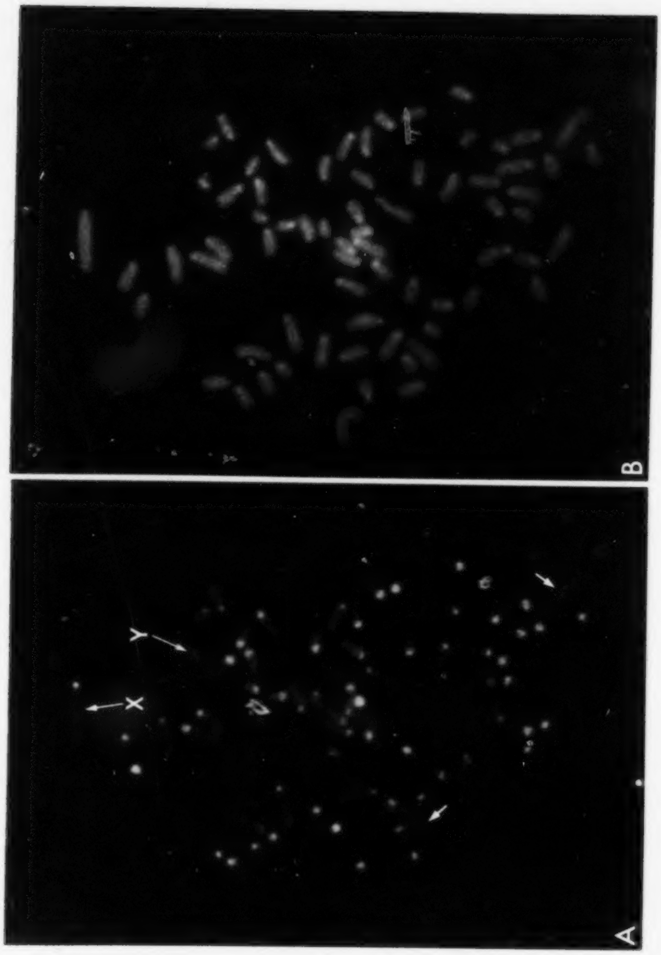

B
A
Y
X

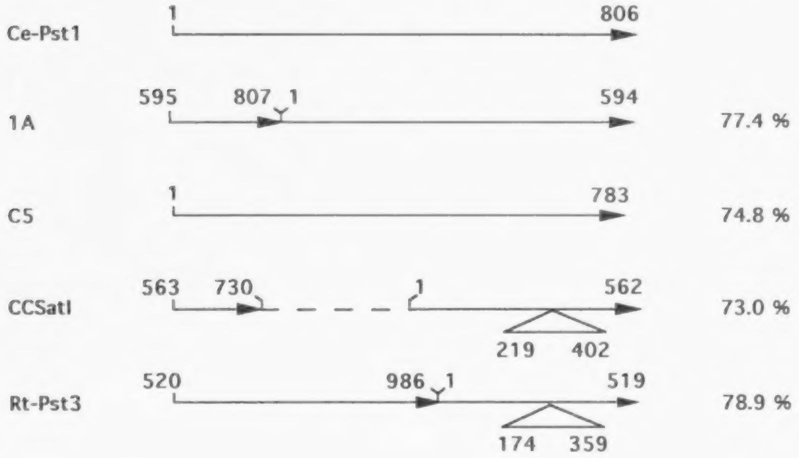
Figure 4.3. DNA sequence and comparison strategy for the Ce-Pst1 clone.

(**A**) The complete 806 bp sequence of the Ce-Pst1 DNA clone is shown with a stretch of 8 continuous cytosine residues indicated in a box. Fifteen residues of CAGG or GAGG tetranucleotides are underlined. (**B**) Strategic alignment of the Ce-Pst1 clone with centromeric satellite DNA sequences from the Indian muntjac (1A), Chinese muntjac (C5), European roe deer (CCSatI), and Canadian woodland caribou (Rt-Pst3). Certain nucleotide positions of each monomer are denoted and a 260 bp gap in the CCSatI sequence is shown as a dotted line. The percent sequence homology of each deer clone to the Ce-Pst1 sequence is indicated to the far right.

A

```
  1 CTGCAGTGAG  TTGAATGCAG  GCTCGTCTTG  CATCTCCCAA  GACGAAAGGA
 51 TGTCTGAATC  CCCTGTGGAG  ACCACAGAGN  AAAGACCTAG  TTCTCCACCT
101 CATCGCGACC  GGAGGCCTCA  CATCCTTTGA  AAACTCCAGA  GGTACGCGGA
151 GATCAGTGTC  TCCAAAGGAG  ACGATGCCTG  ACTCCTCTTG  AAACTTGATA
201 GGAGTCCCAG  GATTCCTGTG  GCACGTGGAA  AGGGACCCTT  GGTCTCCCGC
251 CTCAGCTGGA  GAGGCGTCCC  AATTGCCCTG  CCAAGCCTCG  AGGAGAATCC
301 CGAGCTGTCC  CTCGCAACTA  GGCAGGAGTC  CTGACGTCAC  TGAACAAAAC
351 ACGTGGNTGA  AAGGGCCATC  CCCGTCGGAA  CTCGAGAAAA  TATCCCCAGG
401 TTCCCCCCCC  AACTCGAGAA  AAACCATTGA  GACTTCCCCT  TCNCCGCGAG
451 ATGAGGCCCG  ATTCCCCTGA  CTGCGTGCAG  AGCCATTCAG  TGTTGCACAT
501 CACACATGAA  AGGAGCCTTG  ATTTCCATGA  TGGCACTCCA  GAGAAACCCC
551 AAGAACACTG  TTTCCAGGCT  AGAGGGATCC  TGAGGTCACT  GTAGCAACAC
601 GAAAGAGCTC  CGTGGACCAA  AAATCAACTC  GAGATGAGAG  GTTAGTCCCT
651 GGCTTCGACT  CCAGAGGAAT  ACCACCTTAC  CACAAGCACC  TCAAGAGGAG
701 GCTTCTCTCA  GCTCTAGCTA  TGTGGGAGGG  ACCCTGAGTT  TGCAGCCTCA
751 AGTGGAATGG  ACACCGCGAA  GCCCTGACTC  GAAATAAGGG  TGGATATNCC
801 CTGCAG
```

B



aligned to the previously reported DNA sequences of the 1A, C5, CCSatI, and Rt-Pst3 clones, sequence homologies of 77.4%, 74.8%, 64.2%, and 73.3% were observed, respectively. A maximum sequence similarity of 73.0% was achieved with the CCSatI clone when a 260 bp gap was introduced after nucleotide 730 and a 184 bp region (nucleotides 219 - 402) removed from the CCSatI clone (Figure 4.3B). Likewise, a maximum sequence similarity of 78.9% could be obtained between the Ce-Pst1 and Rt-Pst3 clones when nucleotides 174 - 359 of the Rt-Pst3 clone were removed prior to sequence comparison. It was also noted that nucleotides 174 - 359 of the Rt-Pst3 clone shared 72.6% sequence similarity to the preceeding 184 bp of the same clone (i.e. nucleotides 973 - 985 + 1 - 173).

An internal 31 bp periodicity for each deer monomer clone.

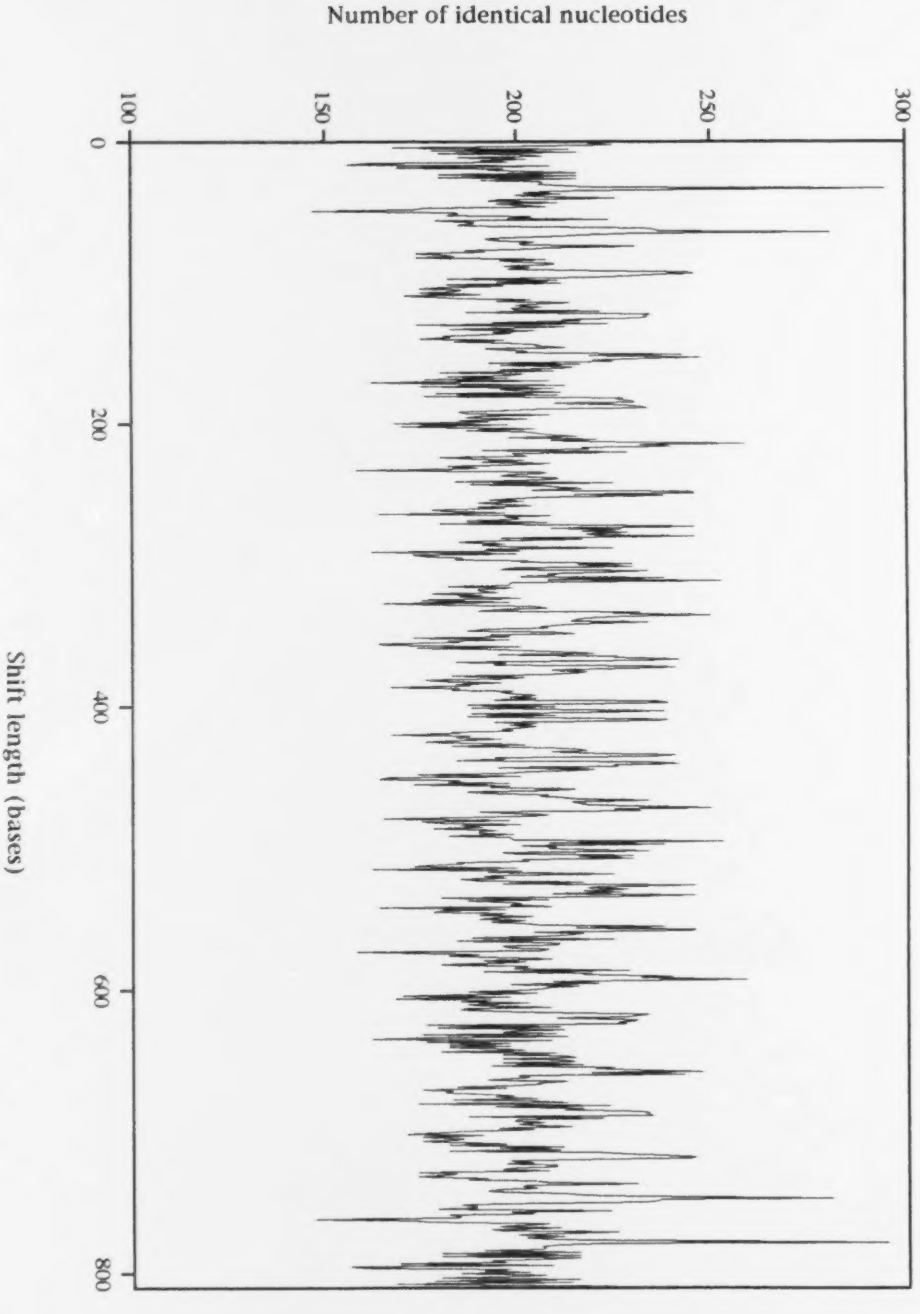
Using the afore mentioned nucleotide shift self comparison system (see Materials and Methods), a total of 805 shifts were made for the Ce-Pst1 clone, 782 shifts for the C5 clone, 727 shifts for the CCSatI clone, and 990 shifts for the Rt-Pst3 clone. Line graphs were produced with the data from nucleotide shift self-comparisons of each deer clone. Each graph exhibited an "in-frame" peak after approximately the first 31 single base shifts. Additional "in-frame" peaks, of varying heights, were then successively observed after approximately every 31 shifts (Figure 4.4A). The proportion of size increase for each "in-frame" peak was calculated for all four cervid monomers, and found to range from 16.4% to 46.8% in the Ce-Pst1 clone, from 9.6% to 33.8% in the C5 clone, from 9.4% to 32.0% in the CCSatI clone, and from 13.4% to 43.1% in the Rt-Pst3 clone (Table 4.1). Consensus sequences for the 31 bp subrepeats in each of the four monomer clones could also be derived (Figure 4.5). These results confirmed the existence of 31 bp subrepeats within all cervid monomer clones studied.

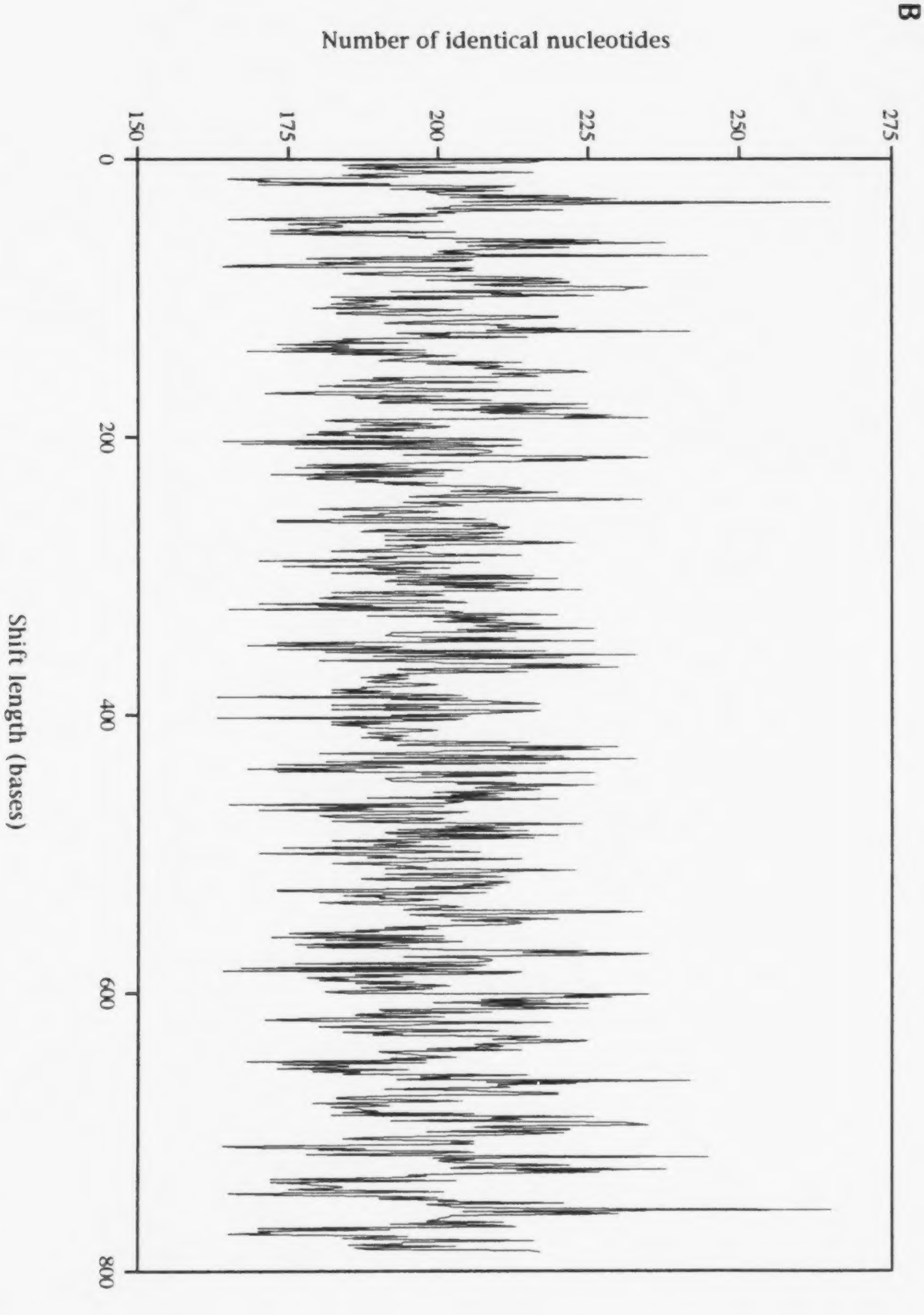
Each monomer-specific 31 bp cervid consensus sequence was found to be extremely similar to the bovine 31 bp subrepeat consensus sequence (Figure 4.6A).

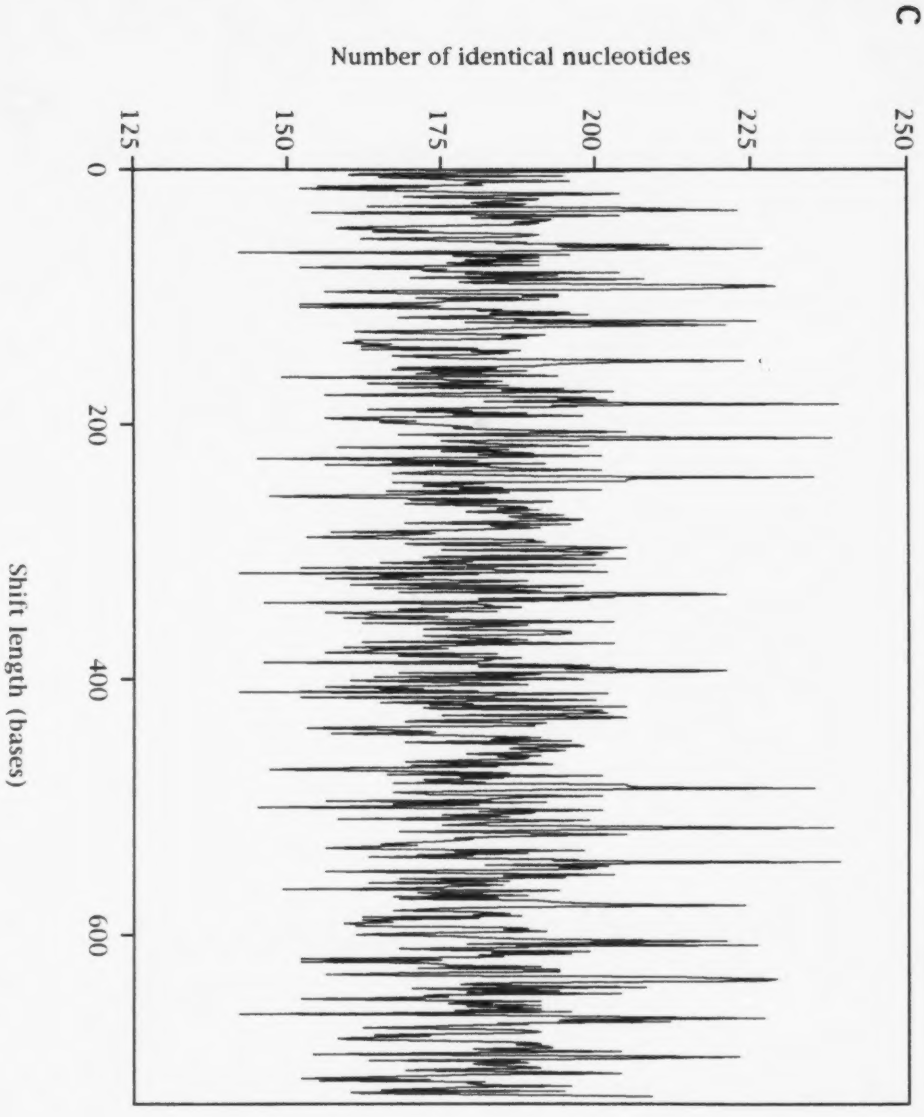
Figure 4.4. The existence of a 31 bp periodicity in cervid centromeric satellite DNA clones: Ce-Pst1, C5, CCSatI, and Rt-Pst3.

Line graphs demonstrating increased DNA sequence similarity (peak) in 31 base shift periodicities during nucleotide shift self-comparisons of clone (**A**) Ce-Pst1, (**B**) C5, (**C**) CCSatI, (**D**) Rt-Pst3.

A


Number of identical nucleotides
100
150
200
250
300
0
200
400
600
800
Shift length (bases)


B
Number of identical nucleotides
150
175
200
225
250
275
0
200
400
600
800
Shift length (bases)


C
Number of identical nucleotides
125
150
175
200
225
250
0
200
400
600
Shift length (bases)

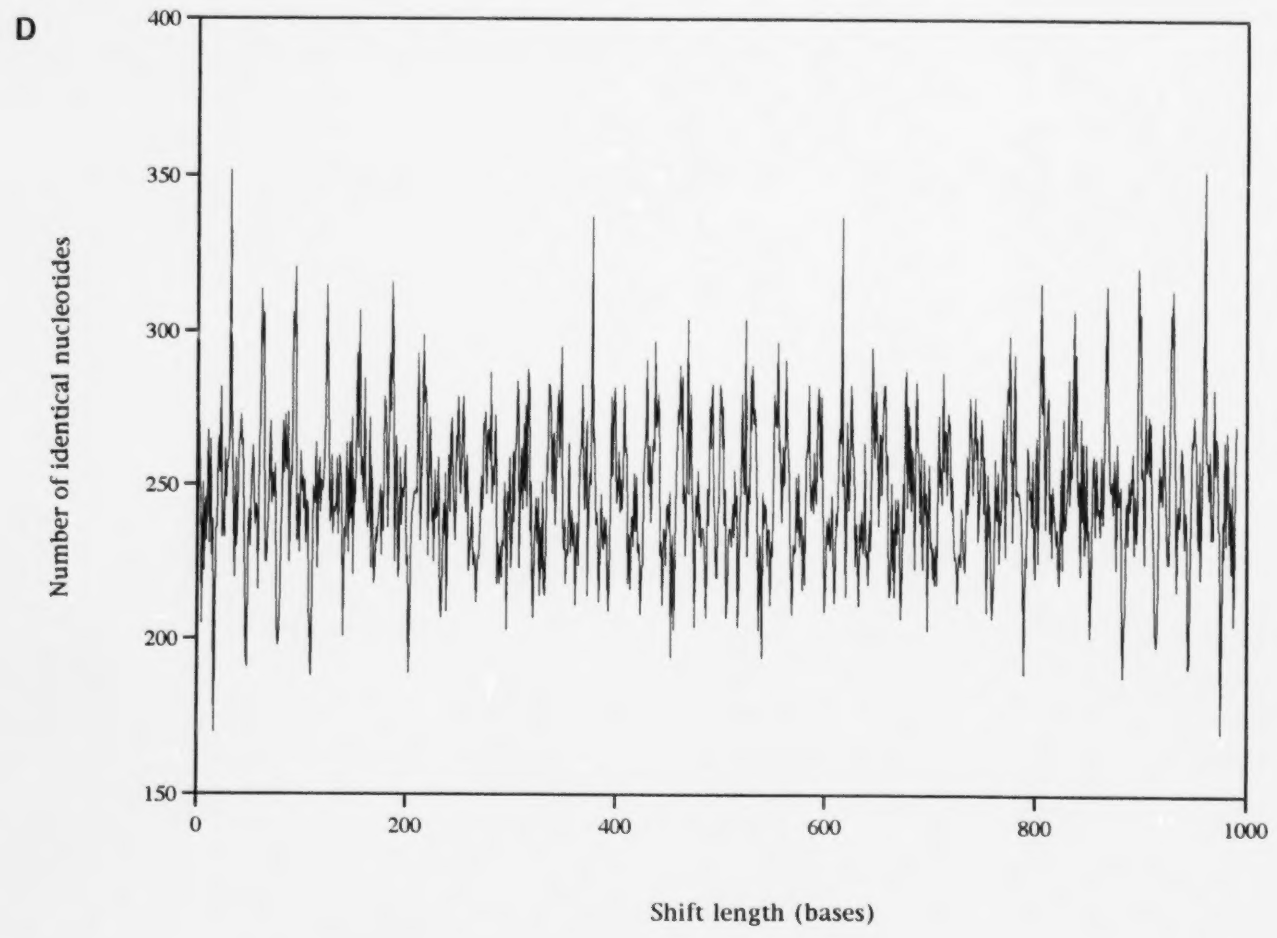

D
Number of identical nucleotides
400
350
300
250
200
150
0
200
400
600
800
1000
Shift length (bases)

TABLE 4.1: SIZE INCREASES OF IN-FRAME PEAKS

Species	Clone	Largest In-Frame Peak Size*	Smallest In-Frame Peak Size*	ANIN	Increase Range of In-Frame Peaks
C. elaphus h.	Ce-Pst1	295	234	201	16.4% - 46.8%
M. reevesi	C5	265	217	198	9.6% - 33.8%
C. capreolus c.	CCSatI	239	198	181	9.4% - 32.0%
R. tarandus c.	Rt-Pst3	352	279	246	13.4% - 43.1%
Averages:					12.2% - 38.9%

Legend:

* - expressed in terms of number of identical nucleotides

ANIN = Average number of identical nucleotides

Increase = (Peak Size - ANIN)/ANIN x 100 %

Figure 4.5. Derivation of consensus sequences for the 31 bp subrepeats in clones Ce-Pst1, C5, CCSatI, and Rt-Pst3.

Derivation of consensus sequences from strategically aligned subrepeats within the (**A**) Ce-Pst1 clone, (**B**) C5 clone, (**C**) CCSatI clone, (**D**) and Rt-Pst3 clone. The nucleotide position of the beginning of each subrepeat is indicated to the left. Deletions in each subrepeat are replaced by *dashes*. The positions where nucleotides have been removed from each subrepeat are indicated by small vertical arrows. Subtracted nucleotides are listed to the right of each subrepeat and separated by commas. The percent sequence similarity of each subrepeat to the consensus sequence is indicated to the far right. In *C*, nucleotide position 8 of the consensus sequence finds the cytosine nucleotide occurring as equally frequent as the adenine nucleotide. Thus, lower case letters are used to represent cytosine and adenine nucleotides at this position in certain subrepeats.

A

```
                                                          Subtracted
Consensus:TCCCTGCCTCAACTCGAGAGGAATCCCGAGT   Nucleotides
        1      ....AGTG.G.T..AT.C.GG.T..TC.                          50.0 %


       29      .G.A.CT.C...GA...A....TGT.T..A.                       58.1 %
       60      C.....TGGAG..CAC....N..AG↓.T...      A                56.3 %
       92      ..T.CA.....T.G...CC...GG..TC.CA                       58.1 %
      123      ...T.↓A--A.....C.....T.CG.G...A      T                62.5 %
      153      ..AG..T...C.AAG....C..TG..T..C.                       61.3 %
      184      --...-.T.G..↓.T..T....G....AG.A      A                62.5 %
      213      .T....TGG..CG.G..A...G.C..TTG..                       58.1 %
      244      ↓....-......G..G......CG....A.T.     C                71.9 %
      275      G......--...↓......GA..........C↓    GC,TG            71.4 %
      308      ......--.G.....A.GC....G...T..↓..↓   C,C              69.7 %
      339      A--...A↓.AA...A..T.G↓GA..GGG.C.-.    A,NT             44.1 %
      370      -...C.T.GG..........AA.↓.....C..G    T                68.8 %
      401      .T..CC..C..........AA..C.AT↓..A      T                65.6 %
      433      ↓....--.T..NC.G.....T..GG.....-.     CT               63.6 %
      463      ....↓..A..GCGTG.-....CC..T.--...     C                59.4 %
      492      G--T...↓↓...↓.AT..A....GC.TT..T.     A,A,C            55.9 %
      524      ...A..ATGGC....C....A..C...A..A                       64.5 %
      555      A.A...TT..C.GG.T.....G....T...G                       64.5 %
      585      ..A...TAG....A...A..-.GCT...T.G                       61.3 %
      615      A..AAAAA...........T..GAGGTT...                       58.1 %
      646      -.....G..↓.G....C.........A..ACC.    T                71.9 %
      678      .A..ACAAG..C...A......GG.TTCTC.                       51.6 %
      709      C↓T..AG..-.TG.-.G....G.C..T....      AG               57.6 %
      740      .TG.A.......G.G..AT...↓.C.G...A-     C                62.5 %
      771      G.....A...↓..A.A↓G.T...↓..N..CT.C    G,A,T            61.8 %
      805      AG
                                                                   ______
                                                    Average:         61.2 %
```

B

Consensus:TCCCTGCCTCAACTCGAGAGGAATCCTGAGT

		Subtracted Nucleotides	
1	C.		
3	..AG..A.↓.↓..-...T.C..G.-...-.	T,GTG	62.9 %
35	..ATCT-.A...GATA.-...G↓.G....AG	T	50.0 %
65	C.....GGGAG..A.T....A..G..CT..C		58.1 %
96	.........T.T..A......CGC↓.CC.-.	G,T	66.7 %
128	CAAT..GAG-.G.........G.A↓.G....	A	59.4 %
160	..AA..AT..C.AAG....C..TG.....C.		61.3 %
192	C.T...G--A..A.G..C........C↓.-.	AG	60.6 %
220	-...A.TGG....AG..T...G..AA..G..		61.3 %
250	-.TG↓......C..T↓...A.TT↓...AT.G	C,T,T	58.8 %
279	-....--T...↓.C....GA..C...C...G↓	GC,G	61.8 %
313	.....--.A....AA-.C....G......CC		67.7 %
347	..G...A↓A..C.C..T.TT↓..GGGCC.-.	A,T	51.5 %
372	C....-.GTT.........AT↓.A..A.T-.	T	62.5 %
404	-..--...G.........AA..-.G.....		74.2 %
431	CT..AC....GC.A.....TA.GG..C..T.		58.1 %
462	C......A-..G.AT↓....C...TGC.T..	C	62.5 %
493	....A--A...↓.C..GA.A...G..↓..T.	A,T	63.6 %
524	...T↓..-...---......C.GA..-.---	GA	57.6 %
549	.T..A.A↓.GG.TC.T.A..TC.-.T.C.--	ATG	44.1 %
580	.↓..A.AAA..G..↓..T.T--.T.T.AA.A-	AA,T	47.1 %
611	---..T.-..G↓..-.....↓..T.T.CTG.↓	G,AT,CT	52.8 %
642	----------...A..A........GACA.C.		48.4 %
665	↓...↓.AAG.A.C..AA...↓...CT..TC.↓..C	G,A,G,C	55.9 %
699	.--G.A.G..TGG.T...↓...↓...T-..C.G	GG,TTG	50.0 %
732	C↓..↓..GA↓.↓...C.....T.--C.....C.	T,AG,A,G	61.1 %
766	---------CG..A..---...---..G..A.↓	T	37.5 %
		Average:	57.5 %

C

Consensus: $\mathrm{TCCCTGC}\genfrac{}{}{0pt}{}{\mathrm{C}}{\mathrm{A}}\mathrm{TCAACTCGAGAGGAACCCCGAGT}$

		Subtracted Nucleotides	
1	.GG.----.....C------------		26.9 %
11	..GG.A.c↓.--..TT..-..TT.--...T.	CGGGGA	51.4 %
43	C......a.TGCG.GC....C..TT..TT..		61.3 %
74	.G..A--a...↓AG.TA....G..TT..T.	A	59.4 %
104	.G.T..ATGG..G.A..↓..TCC...A..AA	TT	51.5 %
137	C....--cA...G..T...A....A.T..↓.	T	71.9 %
167	-.A.C..aGG..TGT..A.T-.T.T.G.G..		51.6 %
196	AG..CAAa.......T...TT..G...A.T.		64.5 %
227	C......a.TGG...C...AC..↓..T..-.	AT	66.7 %
259	C...AT.a.AC..↓.A..T..c↓.TT.--T.	G,TT	52.9 %
292	.A↓↓.A.cGA..T..CC↓..--.....--.C-	GG,AA,T	52.8 %
323	---G.-.T...GG.TT.T......A..A.-.		54.8 %
349	.T.AGA.aAG...A....-.A..G.T.C.↓A	T	56.3 %
381	↓..A.↓.---.T...A...A..GAG.T..T.	C,A	57.6 %
410	C..AG.AT..T.T.GA.C...T.TG..A.C.		51.6 %
441	.TAAACAaG......A......G..↓.--C.	TT	57.8 %
472	CA..↓AGG.-.TG.-......G....T....	AA	60.6 %
503	.TT....c....G.G...T..↓..GA..-.	T	71.9 %
534	......Ac.A↓.A.AAG.T↓..TTT..C..-	G,C	57.6 %
566	AAGAG..T.G..TG.---...CT.GT.TTTC		35.5 %
594	AT.-..Ac--..GAT..AG...TGT.T..A.		48.3 %
622	C.T...TGGAG..C.T....A..A..T.GT.		54.8 %
653	↓..--..c...T..G..C.T.GG..↓AC.TC	C,TT	55.9 %
685	..T-..AcA.--.....C...CT.ATG....		61.3 %
713	..AA...T....----..TT.G		54.5 %
		Average:	55.6 %

D

```
                                                  Subtracted
Consensus:TCCCTGCCTCAACTCGAGAGGAACCCCGAGT         Nucleotides

   1        ....A↓GC.TG.-....------.T.-.          G                    50.0 %
  22      ...G.↓.↓...A-.A....↓N.T------           AT,CAT,TC            50.0 %
  52      .T↓.TG..T..-.G..TC.AG..AG-----          GA                   42.4 %
  81      ....--.AAG...A.↓TC.CC.GT.TA..↓G         T,C                  48.5 %
 112      A.....AG↓..C.C.---..C...G.AA..C↓        G,AG                 52.9 %
 143      CT.....↓...........T..GG.....-.         CCAAA                69.4 %
 181      ...↓...A.TG....C....CC↓....--..         C,TT                 67.6 %
 213      ....CAT-.....A...CC..TGG.T..C-.                              58.1 %
 243      ...T.CAGG......C....ATG.......A                              67.7 %
 274      A.A...T.C...G↓.-........A..A.C.         G                    66.7 %
 306      ..AGCA.AG........--...↓T...T..          TG                   63.6 %
 337      A..↓..T...G.G.....CT..TT...TG.C         CAAA                 54.3 %
 375      ↓.T↓...--.....↓C...A.C.N.T.-..G         TAGG,AAGACGTA,T      47.7 %
 416      AGGAG.↓↓....G.--A.T...T.TGTG...G        TT,T                 44.1 %
 450      G↓..↓..TG.--...A..G....TATAC..C         A,CGAGTT             47.4 %
 486      GAGAA...CT....T..A....GGN.G..T.                              54.8 %
 517      ...T...AGTGG..T..ATAC.GG.T..TC.                              48.4 %
 548      .T.A.--C..↓.↓-..AG...T.T.T..A.          AC,A                 55.9 %
 579      C.....TGGAG..CTT....A..↓..T-..G         G                    56.3 %
 610      ..N........TT.T.N.-....T↓.TC.C↓         G,A                  63.6 %
 642      C↓↓..A--..C.........C..G.G...G          N,G                  66.7 %
 673      .T.A↓..T.N..AAG.T..C..TG..T..C.         T                    56.3 %
 705      --...-.T.G..TA↓..T.....↓...A.A↓         T,T,A                61.8 %
 736      -.....TGG.T...G..A...GT..TT.G..                              61.3 %
 766      G..........C.......A.CGT...TTTG                              67.7 %
 797      G......--..↓....↓..-.TT..T...↓↓         GC,A,G,G             66.7 %
 830      .....-..GTT...A..C..-.GT..T..↓.         C                    65.6 %
 861      -.A...AAGA......T.TT↓..GGG.C.-.         TT                   51.5 %
 892      -...C..G.-..G......AC↓.....AG..         T                    68.7 %
 922      -...C...G..........AA....AT...A                              74.2 %
 952      ↓....--...AGNGG.....T..GG...C.T.        CT                   57.6 %
 983      C......AG                                                    66.7 %
                                                                       ------
                                                  Average:             58.6 %
```

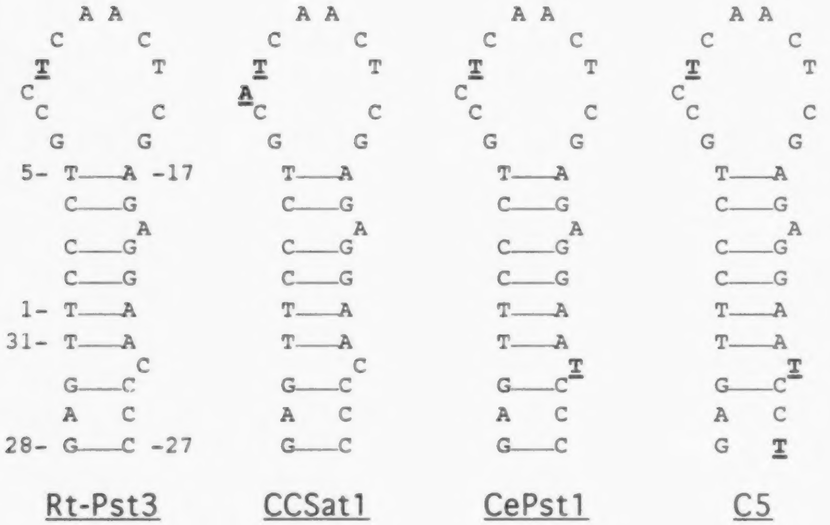
Figure 4.6. DNA sequence analyses of the monomer-specific 31 bp consensus sequences.

(**A**) A comparison of the bovine 31 bp subrepeat to each cervid monomer-specific 31 bp consensus sequence. The bovine 31 bp subrepeat is designated at the top of the figure. The subrepeat consensus sequences for monomers Rt-Pst3, CCSatI, Ce-Pst1, C5, and IA are displayed below the bovine consensus sequence, respectively. Nucleotides which are similar to corresponding bases in the bovine sequence are replaced with *dots*. Horizontal arrows beneath the bovine 31 bp consensus sequence indicate the two pairs of related sequences described by Plucienniczak *et al*. (1982). (**B**) An alternative representation of the 31 bp consensus sequences considering a self complementary region and an inverted repeat. Each 31 bp consensus sequence is represented by a separate hairpin loop. The complementary sections of each consensus sequence are shown base paired to each other by short horizontal lines. Nucleotides which differ from the bovine 31 bp subrepeat consensus sequence are in bold and underlined. Some nucleotide positions are indicated only in the hairpin structure for the Rt-Pst3 clone.

A

```
         1       10        20        31
         |        |         |         |
Bovine:  TCCCTGCCGCAACTCGAGAGGAACCCCGAGT

Rt-Pst3: ........T......................
CCSat1:  ........T......................
                A
Ce-Pst1: ........T..............T.......
C5:      ........T..............T..T....
IA:      ........T..............T..T....
                 .              G
```

B



5-
-17
1-
31-
28-
-27
Rt-Pst3
CCSat1
CePst1
C5

Further examination of each monomer-specific 31 bp consensus sequence for the presence of internal repetitions revealed nucleotides 1 - 4 and 5 - 8 as two related tetranucleotides and nucleotides 9 - 19 and 20 - 31 as another pair of direct, although imperfect, repeats. Alternatively, nucleotides 17 - 27 appeared to be complementary to nucleotides 28 - 31 and 1 - 5, permitting these sequences to also resemble stem loop-like structures (Figure 4.6B).

Discussion.

Ce-Pst1 and a previously isolated 770 bp red deer clone.

A 770 bp repetitive BamHI DNA fragment was previously isolated from the European red deer genome (Lima-de-Faria *et al.* 1986) and used in Southern blot and radioactive *in situ* hybridization experiments. In that communication, the 770 bp cloned DNA fragment produced distinct hybridization bands in a 0.8 kb register with BamHI-, EcoRI-, MspI- and HpaII- digested red deer genomic DNAs, similar to the Southern blot results in the present study. As well, *in situ* hybridization experiments showed hybridization signals in the centromeric region of all autosomes (with the exception of a pair of metacentric autosomes), as was observed with the Ce-Pst1 clone in the present study. Overall, these findings would seem to suggest that the 770 bp repetitive DNA fragment obtained by Lima-de-Faria and collaborators could belong to the same satellite DNA family as the 806 bp Ce-Pst1 clone. However, the 770 bp fragment also hybridized to the centromeric region of the submetacentric Y-chromosome and not to the acrocentric X-chromosome (Lima-de-Faria *et al.* 1986). In the present study, FISH experiments localized the Ce-Pst1 clone to the centromeric region of the X but not the Y-chromosome. If the 770 bp repetitive DNA fragment and the Ce-Pst1 clone are both derived from the same DNA family, it is unknown whether certain sequences are present in the 770 bp fragment which allowed it to recognize certain centromeric DNA sequences specific for the red deer Y-chromosome. Likewise, other DNA sequences may have been present in the Ce-Pst1 clone which detected red deer X-chromosome specific centromeric DNA sequences. However, since DNA sequence data is unavailable for the 770 bp DNA fragment, such speculations cannot be tested. Alternatively, these two DNA fragments may be derived from different centromeric satellite DNA families which both happen to consist of 0.8 kb tandemly organized monomers but exhibit slightly different chromosomal localization patterns. It is noteworthy that other FISH studies using major cervid centromeric satellite DNA clones as probes have so far consistently shown hybridization to the centromeric region of deer acrocentric X-chromosomes (e.g. Bogenberger *et al*. 1987, Lin *et al*. 1991).

CAGG and GAGG tetramers in the Ce-Pst1 DNA sequence.

The CAGG and GAGG tetramers observed in the Ce-Pst1 clone did not appear to be arranged in any regular distance from one another, suggesting that its frequent occurrence is probably not due to the presence of internal 31 bp subrepeats. An increased appearance of these motifs has previously been reported in the centromeric satellite DNA monomers of other deer species (Bogenberger *et al*. 1987, Scherthan 1991, Lee *et al.* 1994), and can be observed in human gamma 8 and X centromeric satellite DNA sequences (Lin *et al.* 1993, Lee *et al.* 1995) as well as certain human minisatellites (Jeffreys *et al.* 1985). These motifs are thought to possibly play a role in the constitution of recombination hot spots (Steinmetz *et al*. 1986, Krowczynska *et al*. 1990), and is also associated with DNA deletion hot spots in certain human gene loci (Huff *et al.* 1995). If indeed these motifs specify the occurrence of recombination events, they may suggest a role for certain centromeric DNA sequences in karyotypic evolution. Scherthan (1990) proposed that the ends of acrocentric chromosomes could be brought in close proximity clustering centromeric and telomeric DNA during the bouquet stage of meiosis. Breakages and reunions at centromeric and/or telomeric DNAs could then promote tandem fusions of these acrocentric chromosomes, dictating the karyotypic evolution necessary to produce karyotypes like those of the present day Indian muntjac (Hsu *et al.* 1975).

31 bp subepeats in the cervid monomers.

Although the 31 bp subrepeat consensus sequences for different monomer clones studied were found to be quite similar, substantial sequence divergence could be seen between the individual subrepeats of each cervid centromeric satellite DNA monomer (Figures 4.5). Therefore, it is not surprising that an internal periodicity could not be detected in several cloned cervid centromeric DNA monomers without the use of a sequence comparison strategy aimed at detecting internal direct repeats.

Each monomer-specific 31 bp cervid consensus sequence could also be represented as a hairpin-like structure (Figure 4.6C). Although the thermodynamic stability of a hairpin structure composed of a single 31 bp subrepeat could not be established following suggestions proposed by Tinoco *et al.* (1973), thermodynamically stable hairpins or foldback structures have already been identified in heterochromatin-associated satellite DNAs of rodents (Modi 1993), humans (Grady *et al.* 1992, Catasti *et al.* 1994), and other higher eukaryote species (Ferrer *et al.* 1995).

Evolution of cervid centromeric satellite DNA.

To date, only one family of centromeric satellite DNA has been identified in cervid species. This major cervid centromeric satellite DNA family has been shown to consist of repeat units which vary in size among various deer species. Bogenberger *et al.* (1987) demonstrated that satellite DNAs, homologous to the Indian muntjac 1A satellite DNA clone, were organized into 0.8 kb monomer units in deer species belonging to the paleontological division, plesiometacarpalia. Deer species belonging to the telemetacarpalia division were shown to have the majority of their satellite 1A homologous DNA organized into monomer units of about 1 kb.

Taking into account the existence of internal 31 bp subrepeats in all cervid centromeric satellite DNA clones studied, various number of amplifications of the ancestral 31 bp DNA sequence could have occurred to result in the various monomer sizes within this satellite DNA family. For example, 25 amplifications of the 31 bp subrepeat could account for the approximate 781 bp B1 and C5 monomer clones from the Indian and Chinese muntjacs, respectively. One further subrepeat amplification could account for the 806 bp monomers of the 1A and Ce-Pst1 clones from the Indian muntjac and red deer, respectively. Similarly, 32 amplifications of the subrepeat could have generated the 991 bp Rt-Pst3 monomer clone of the caribou.

When the Rt-Pst3 clone of the telemetacarpalial caribou was sequenced (Lee *et al.* 1994), the first 191 bp of the 991 bp clone was shown to share approximately 60% DNA sequence similarity to the last 191 bp of the same clone. This amount of sequence similarity seemed significant enough to warrant the postulation that these 991 bp monomers could have resulted from selective amplification of an ancestral 806 bp DNA monomer along with approximately 185 bp of an adjacent DNA monomer. Further sequence analyses from the present study have shown that juxtaposed 184 bp regions within the Rt-Pst3 clone share 72.6 % sequence similarity. Due to this observed high sequence similarity between two adjacent 184 bp regions, it seems more favorable to postulate that these 991 bp monomers arose from an approximate 184 bp duplication within the original 806 bp monomer.

Simpson (1984) identified the Bovidaes, Cervidaes, and Tragulidaes as separate families of the suborder Ruminantia. Bovids are believed to have first appeared in the early Miocene of Africa and Eurasia while cervids are thought to have first appeared in the early Miocene of Eurasia (Scott and Janis 1987). The divergence of these two families from a common ancestor is estimated at about 25 million years ago (Irwin and Wilson 1990). The presence of a 31 bp subrepeat in both cervid and bovid centromeric satellite DNA monomers substantiates the idea that this sequence was present in the genomes of a common ancestor, thought by some to resemble the Tragulidae species (Scott and Janis 1987).

References:

Bogenberger JM, Neumaier PS, Fittler F. (1985) The Muntjak satellite IA sequence is composed of 31-bp-pair internal repeats that are highly homologous to the 31-base-pair subrepeats of the bovine satellite 1.715. *Eur J Biochem* 148: 55-59.

Bogenberger JM, Neitzel H, Fittler F. (1987) A highly repetitive DNA component common to all cervidae: its organization and chromosomal distribution during evolution. *Chromosoma* 95: 154-161.

Catasti P, Gupta G, Garcia AE, Ratliff R, Hong L, Yau P, Moyzis RK, Bradbury EM. (1994) Unusual structures of the tandem repetitive DNA sequences located at human centromers. *Biochemistry* 33: 3819-3831.

Ferrer N, Azorin F, Villasante A, Gutierrez C, Abad JP. (1995) Centromeric dodeca-satellite DNA sequences form fold-back structures. *J Mol Biol* 245: 8-21.

Gaillard C, Doly J, Cortadas J, Bernardi G. (1981) The primary structure of bovine satellite 1.715. *Nucleic Acids Res* 9: 6069-6082.

Grady DL, Ratliff RL, Robinson DL, McCanlies EC, Meyne J, Moyzis RK. (1992) Highly conserved repetitive DNA sequences are present at human centromeres. *Proc Natl Acad Sci U.S.A.* 89: 1695-1699.

Gustavsson I, Sundt CO. (1968) Karyotypes in five species of deer (*Alces alces L., Capreolus capreolus L., Cervus elaphus L., Cervus nippon temm. and Dama dama L.). Hereditas* 60: 233-247.

Horz W, Altenburger W. (1981) Nucleotide sequence of mouse satellite DNA. *Nucleic Acids Res* 9: 683-696.

Hsu TC, Pathak S, Chen TR. (1975) The possibility of latent centromeres and a proposed nomenclature system for total and whole arm translocations. *Cytogenet Cell Genet* 15: 41-49.

Huff V, Jaffe N, Saunders GF, Strong LC, Villalba F, Ruteshouser EC. (1995) WT1 Exon I deletion / insertion mutations in Wilm's tumor patients, associated with di- and trinucleotide repeats and deletion hotspot consensus sequences. *Am J Hum Genet* 56: 84-90.

Irwin DM, Wilson AC. (1990) Concerted evolution of ruminant stomach lysozymes. *J Biol Chem* 265: 4944-4952.

Jeffreys AJ, Wilson V, Thein S. (1985) Hypervariable minisatellite regions in human DNA. *Nature* 314: 67-73.

Jobse C, Buntjer JB, Haagsma N, Breukelman HJ, Beintema JJ, Lenstra JA. (1995) Evolution and recombination of bovine DNA repeats. *J Mol Evol* 41: 277-283.

Krowczynska AM, Rudders RA, Krontiris TG. (1990) The human minisatellite consensus at breakpoints of oncogene translocations. *Nucleic Acids Res* 18: 1121-1127.

Lee C, Ritchie DBC, Lin CC. (1994) A tandemly repetitive, centromeric DNA sequence from the Canadian woodland caribou (*Rangifer tarandus caribou*): its conservation and evolution in several deer species. *Chromosome Res* 2: 293-306.

Lee C, Li X, Jabs EW, Court D, Lin CC. (1995) Human gamma X satellite DNA: an X chromosome specific centromeric DNA sequence. *Chromosoma* 104: 103-112.

Lima-de-Faria A, Arnason U, Widegren B, Essen-Moller J, Isaksson M, Olsson E, Jaworska H. (1984) Conservation of repetitive DNA sequences in deer species studied by southern blot transfer. *J Mol Evol* 20: 17-24.

Lima-de-Faria A, Arnason U, Widegren B, Isaksson M, Essen-Moller J, Jaworska H. (1986) DNA cloning and hybridization in deer species supporting the chromosome field theory. *Biosystems* 19: 185-212.

Lin CC, Sasi R, Fan Y-S, Chen Z-Q. (1991) New evidence for tandem chromosome fusions in the karyotypic evolution of Asian muntjacs. *Chromosoma* 101: 19-24.

Lin CC, Sasi R, Lee C, Fan YS, Court D. (1993) Isolation and identification of a novel tandemly repeated DNA sequence in the centromeric region of human chromosome 8. *Chromosoma* 102: 333-339.

Modi WS. (1993) Rapid, localized amplification of a unique satellite DNA family in the rodent *Microtus chrotorrhinus*. *Chromosoma* 102: 484-490.

Nowak R. (1994) Mining treasures from 'junk DNA'. *Science* 263: 608-610.

Pech M, Streeck RE, Zachau HG. (1979) Patchwork structure of a bovine satellite DNA. *Cell* 18: 883-893.

Plucienniczak A, Skowronski J, Jaworski J. (1982) Nucleotide sequence of bovine 1.715 satellite DNA and its relation to other bovine satellite sequences. *J Mol Biol* 158: 293-304.

Scherthan H. (1990) The localization of the repetitive telomeric sequence $(TTAGGG)_n$ in two Muntjac species and implications for their karyotypic evolution. *Cytogenet Cell Genet* 53: 115-117.

Scherthan H. (1991) Characterization of a tandem repetitive sequence cloned from the deer *Capreolus capreolus* and its chromosomal localisation in two muntjac species. *Hereditas* 115: 43-49.

Scott KM, Janis CM. (1987) *Phylogenetic relationships of the Cervidae, and the case for a Superfamily "Cervoidea"*. In: Wemmer CM, ed. Biology and Management of the Cervidae. Washington, D.C., Smithsonian Institute Press. pp. 3-20.

Simpson CD. (1984) *Artiodactyls*. In: Anderson S, Jones JK, eds. Orders and Families of Recent Mammals of the World. New York, John Wiley and Sons. pp. 563-587.

Singer MF. (1982) Highly repeated sequences in mammalian genomes. *Intern Rev Cytol* 76: 67-112.

Steinmetz M, Stephan D, Lindahl FK. (1986) Gene organization and recombinational hotspots in the murine major histocompatibility complex. *Cell* 44: 895-904.

Taparowsky EJ, Gerbi SA. (1982) Sequence analysis of bovine satellite I DNA (1.715 gm/cm^3). *Nucleic Acids Res* 10: 1271-1281.

Tinoco Jr I, Borer PN, Dengler B, Levin MD, Uhlenbeck OC, Crothers DM, Bralla J. (1973) Improved estimation of secondary structure in ribonucleic acids. *Nature New Biology* 246: 40-41.

Yu LC, Lowensteiner D, Wong EFK, Sawada I, Mazrimas J, Schmid C. (1986) Localization and characterization of recombinant DNA clones derived from the highly repetitive DNA sequences in the Indian muntjac cells: Their presence in the Chinese muntjac. *Chromosoma* 93: 521-528.

CHAPTER 5

ORGANIZATION AND EVOLUTION OF CERVID SATELLITE I DNA

A version of this chapter has been accepted for publication:

Lee C, Court DR, Cho C, Haslett JL, Lin CC. (1996) Higher-order organization of subrepeats and the evolution of cervid satellite I DNA. *Journal of Molecular Evolution* (In Press).

Prologue

As a result of the drastic karyotypic evolution in the family Cervidae, certain deer species offer an excellent system for investigating mammalian centromere structure and function. To date, only one cervid centromeric satellite DNA family has been identified and was found to be conserved in all deer species studied. This DNA family is organized as 0.8 kb repeat units in plesiometacarpalial deer and 1 kb repeat units in telemetacarpalial deer. The previous chapter reported the presence of similar 31 bp subrepeats in both 0.8 kb and 1 kb monomer clones. However, the existence of these two different sized monomers in cervid satellite I DNA still remains a contradiction to the definition of a satellite DNA family. The following chapter examines the intragenomic and interspecific conservation of this DNA family and provides critical data for explaining the genesis, organization, and evolution of this mammalian centromeric satellite DNA family.

Introduction

Mammalian centromeric DNAs studied to date consist of tandemly organized, repetitive DNA sequences (i.e. satellite DNA). Although more than one satellite DNA family may exist in the centromeres of a given mammalian species, usually a single centromeric satellite DNA family is predominant, often accounting for 5% or more of a species' genome. Each centromeric satellite DNA family is characterized by repeat units (monomers) of relatively consistent length. For example, primate alphoid DNA consists of 171 bp monomers (Rosenberg *et al.* 1978) and mouse major satellite DNA is defined by 234 bp monomers (Horz and Altenburger 1981). Although substantial nucleotide sequence variations are commonly observed amongst monomers of a satellite DNA family, these monomers can be organized in a hierarchical fashion into higher-order repeats which have near identical sequences (reviewed in Willard and Waye 1987).

In the family cervidae, the prominent centromeric satellite DNA family can be referred to as either major cervid centromeric satellite DNA or cervid satellite I DNA (because of its homology to bovine satellite I DNA) and is localized to the centromeric region of nearly all cervid acrocentric chromosomes. Interestingly, monomers of this DNA family vary in size between deer species. Those deer belonging to the palaentological division telemetacarpalia (Brooke 1878, Goss 1983) have their major cervid centromeric satellite DNA primarily organized into 1 kb monomers. Other deer, retaining the more proximal remnants of the second and fifth metacarpals (i.e. belonging to the plesiometacarpalia division), have their cervid satellite I DNA primarily organized into 0.8 kb monomers (Bogenberger *et al.* 1987). Higher-order repeats consisting of a number of tandemly arranged cervid satellite I monomers have not been reported.

The presence of two distinctly different sized monomers in this single cervid satellite DNA family is unusual. Previous investigations involving sequence analyses between these two different sized monomers have revealed the presence of 31 bp subrepeats in all monomers examined, and the occurrence of a tandem 186 bp duplication within a single 1 kb monomer (Lee and Lin 1996). In the present study, we demonstrate that the 0.8 kb and 1 kb monomers actually represent higher-order structures of the 31 bp subrepeats. As well, sequence analyses of other 1 kb cervid satellite I DNA monomers consistently revealed a 0.18 kb tandem duplication. These data offer an explanation for the presence of two distinctly different sized monomers in a single satellite DNA family and provide insights into the evolution of this satellite DNA. Furthermore, extremely high intraspecific sequence similarities of cervid satellite I DNA monomers validate the use of a single monomer for each deer species when elucidating phylogenetic relationships.

Materials and methods:

Source of genomic DNAs. Fallow deer (*Dama dama*) genomic DNA was prepared from an established testis cell line (specimen 1) obtained from the University of Ferrara, Ferrara, Italy. Specimens 2, 3, and 4 were purified genomic DNA samples from unknown sexes of wapiti (*Cervus elaphus canadensis*), white tailed deer (*Odocoileus virginianus*), and the North American moose (*Alces alces*), respectively, obtained from the Bovine Blood Typing Lab, Saskatchewan Research Council, Saskatoon, Canada. Mule deer (*Odocoileus hemionus*) genomic DNAs were prepared from a liver sample of a healthy 2 year old male (specimen 5) and a kidney sample from a second healthy male mule deer (specimen 6).

Cloning and characterization of cervid satellite DNAs. All cervid satellite I DNA clones were obtained from 0.8 kb or 1 kb prominent ethidium bromide-stained bands observed after electrophoretic fractionation of restriction endonuclease digested genomic DNAs. DNA fragments were cloned into pUC19 plasmid vector and propagated in *E. coli* DH5α bacteria. To determine the genomic organization of each clone, ^{32}P-dCTP-labeled clones were used as probes to Southern blots containing endonuclease digested deer genomic

DNAs. Southern blot hybridizations, washings, and exposures were as previously described (Lin *et al.* 1991). Clones were also biotin-labeled and used as probes in FISH experiments to red deer chromosomes. Protocols for chromosome preparations and FISH experiments were as previously described (Lee *et al.* 1994).

Each clone was digested to produce smaller fragments for subcloning and sequencing. All sequencing reactions were performed with dideoxy chain termination Sequenase kits (United States Biochemical Corporation) and read on an ABI DNA Sequencer (Model 373). The DNA sequences of the subclones were combined to produce a complete DNA sequence for each monomer clone. Clones Dd-Pst1, Ce-Msp1, Ov-Msp1, Aa-Msp1, and Oh-Msp1 were designated as representative monomers for the fallow deer, wapiti, white tailed deer, North American moose, and mule deer, respectively.

DNA sequence analyses and comparisons.

Sequence analyses first involved the identification of 31 bp subrepeats in all representative monomer clones, using the method of Plucienniczak *et al.* (1982). These methods are described in detail in another communication (Lee and Lin 1996).

Intragenomic monomer sequence conservation was examined by sequence comparisons between available clones from a white tailed deer, a North American moose, and a mule deer. Since mule deer clones Oh-Msp2 and Oh-Msp3 each contained two complete 1 kb monomers, individual monomers in these clones were distinguished from one another by the suffix designation *a* or *b*.

Intraspecific monomer sequence conservation could only be examined in the mule deer as it was the only species where complete monomers were available from more than one animal. For these comparisons, clone Oh-Msp1 from one animal was compared to each 1 kb monomer in clones Oh-Msp2 and Oh-Msp3 of a second animal. To investigate the extent of subrepeat sequence divergence in a particular species, mule deer clones Oh-Msp1 and Oh-Msp2a were arbitrarily chosen for subrepeat sequence comparisons.

Interspecific monomer sequence conservation of cervid satellite I DNA was determined from all possible pairwise sequence comparisons between the ten available representative monomers (five newly isolated representative monomers from the present study and five previously characterized satellite DNA clones). The previously characterized clones include the Ce-Pst1 clone of the red deer (Lee and Lin 1996), 1A clone of the Indian muntjac (Bogenberger *et al.* 1985), C5 clone of the Chinese muntjac (Lin *et al.* 1991), CCSatI clone of the roe deer (Scherthan 1991), and the Rt-Pst3 clone of the caribou (Lee *et al.* 1994).

Results:

Initial characterization of newly isolated cervid satellite I DNA clones.

Table 5.1 lists the nine newly isolated cervid satellite I DNA clones for which complete DNA sequences were obtained (Figure 5.1). These include one clone from a fallow deer, one clone from a wapiti, two clones from the white tailed deer, two clones from a North American moose, and three clones from two mule deer. Among these clones, seven clones contained single monomers and two clones (Oh-Msp2 and Oh-Msp3), each consisted of two intact, adjacent monomers.

Southern blot hybridizations of each newly isolated clone to XbaI-digested genomic DNAs from the five deer species studied consistently revealed type A-like ladder patterns, similar to Southern blots for other deer species (Lee *et al.* 1994). These ladder patterns of hybridized fragments are characteristic of tandemly arranged repetitive DNA (i.e. satellite DNA). Hybridization fragments were observed in a 0.8 kb register for the fallow deer and wapiti genomic DNAs and a 1 kb register for the white tailed deer, North American moose, and mule deer genomic DNAs.

Table 5.1. Satellite DNA Clones Isolated from 5 Deer Species

Clone	Clone Description	Size	Specimen	Source	GenBank No.
Dd-Pst1	Dama dama PstI Clone 1	804 bp	1	testis cell line	U53515
Ce-Msp1	Cervus elaphus c. MspI Clone 1	680 bp	2	genomic DNA	U53516
Ov-Msp1	Odocoileus virginianus MspI Clone 1	990 bp	3	genomic DNA	U53517
Ov-Msp2	Odocoileus virginianus MspI Clone 2	748 bp	3		
Aa-Msp1	Alces alces MspI Clone 1	987 bp	4	genomic DNA	U53518
Aa-Msp2	Alces alces MspI Clone 2	712 bp	4		
Oh-Msp1	Odocoileus hemionus MspI Clone 1	991 bp	5	liver	U53519
Oh-Msp2	Odocoileus hemionus MspI Clone 2	1970 bp	6	kidney	U55813
Oh-Msp3	Odocoileus hemionus MspI Clone 3	1975 bp	6		U55814

Figure 5.1. Nucleotide sequences of 9 monomer clones from 5 deer species.

The complete DNA sequences of clones (**A**) Dd-Pst1, (**B**) Ce-Msp1, (**C**) Ov-Msp1, (**D**) Ov-Msp2, (**E**) Aa-Msp1, (**F**) Aa-Msp2, (**G**) Oh-Msp1, (**H**) Oh-Msp2, (**I**) Oh-Msp3 are shown.

A. Nucleotide sequence for clone Dd-Pst1:

```
  1   CTGCAGTGTA TACAATGCAG GCTCGTGTTG CATCTCCCAA GACGAAGGGA TGTCTGAATC CCCTGTGGAG
 71   ACCACAGAGA AAGACCTAGT TCTCCGCCTC ATCGCGCCCG GAGGCCTCAC ATTTTTTGAA AACTCGAGAG
141   GTTCGCGGAG ATCCATGCTT CCAAAGGGGA CGATGCCTGA CTCCTCTTGA AAATGGATAG GAGTCCCAGG
211   ATCCCTGTGG CACGTGGAAA GGGACCCTGG GTCTCCCACC TCATGTGGAG ACGCGTCCCA ATTGCCATGC
281   CAAGACTCGA GGAGAATCCC GAGGTGTCCC TCGCACCTAG GCAGGAGTCC TGACGTCGCT GAACAAACAC
351   GTGTCTGGAA GGGCCCTCCC CGCCGTAACT TGAGAATATA CCCCAGGTTC GGGCCGCAAC TCGAGAAAAA
421   CCATGAGACT TTCCCCTCGC CATGAGATGA GGCCTGATTC CCCTGCACTG TGTGTAGAGC AATTCCGTGC
491   TCCACATCCA AAGCTGAAAG GAGCCTTGAT TTCCTTGATG GCACTCCAGA GAAAACCCAA GATCAGTGTC
561   TCAAATCTTG AGGGATCCTT ATGTCACTGT AGCATCACGA AAGGGCTCCG TGGACCCCAA ATCAACTCGA
631   GATGAGAGCT TAGTCCCTAG TTTCGACTCA AGAGGAATAC GAACTTTCCA CAAGCACCTG AAGAGGAGGC
701   TTCTCTCAGC TACAGGCATG TGAGAGGGAC CCTGAGTTTG TGGCCTCAAG TGGAATGGAC ACCGAGATGC
771   TCTGACTCGA AGTAAGGCAG GATATNCCCT GCAG
```

B. Nucleotide sequence for clone Ce-Msp1:

```
  1   CCGGAGCCTC ACATCCTTTG AAAACTCNAG AGTACGCGGA GATCAGTCCT CCAAAGGAGA CGATGCCTGA
 71   CTCCTCGTGA AACTTGATAG GAGTCCCAGG ATTCCTGTGG AACATGGAAA GGGACCCTTG GTCTCCCGTC
141   TCAGGTGGAG AGGCGTCCCA ATTGCCCTGC CAAGCCTCGA GGAAAATCCC GAGTTGTCCC TCGCAACTAG
211   GCAGGAGTCC TGACGTCGCT GAACAAACAT GTNGTTGGAA GGGCCATCCC TGTCGTAACT CGAGAATATA
281   CCCCAGGTTC CCGCCGCAAC TCGAGAAAAA CCATGAGACT TCCCCNTCGC CGCGAGATGA GGCCCGATTC
351   CTCTGCACTG CGTGCAGAGC AATTCCGTGT TGCACATCAC ACATGAAAGG AGTCTTGATT TCCTTGGTGG
421   CACTCCAGAG AAACCGCAAG AACACTGTTT CAAGGCTAGA GGGATCCTGA GGTCACTGTA GCAACACGTA
491   AGAGCTCCGT GGACCCCAAA ACAACTCGAG ATGAGAGGTT AGTCCCTGGA TTCGACTCCA GAGGAATACC
561   ACTTTACCAC AAGCACCTCA AGAGGAGGAT TCTATCAGCT CTAGGTATGT GAGAGGGACC CTGATTTTGC
631   GGCCTCAAGT GGAATGGACA CTGCGATGCC CTGACGCGAA ATAAGGCCGG
```

C. Nucleotide sequence for clone Ov-Msp1:

```
  1   CCGGGTTTCC TTGCAGTGGC TTGAATACAG TCTAGTCTTT CATCTCACAA GGTGAAGGGA TGTCTGCACC
 71   ACCTGTGGAG TCCCTAGAGA AAGCCCTAGG TCCCCGACTC CTCTCGACGG AATGCCTCAC ATGCCTCTGA
141   CACCTCGAGA AGCACGCAGA GCTCACTNTT TCAAAATTTG ACGGAGCCTG ACTCCTCTTG AACATTGACA
211   GGAATCCCAG TATCCCTGTG GCTACTGGAA AGGGACCCTG GGTCTCCCGA CTCACCTCGA GAAGCGTCCC
281   TTTTGCCCTG CCAAGCCTCG AAGAGACTCC TGAGGTGTCC CTCGTAACTC AACAGGACTC CTGACTTCGT
351   TGANCAAACA CGAGGGTTGA AGGGGCATCC CCGCCGTAAA GTCGAGAATG TACCCCAGGT CCGGGCCGGA
421   ACTCGAGAAA AACCATGAGA CTTCCCCCTC CAAGCGAGAT ATGAGGCCCG ATTCCCCTGC ACTGCATGCA
491   TANCAATTCC GTGAGACCCC TGAAACATGT CAGGAGCCTC GACTTGCTTG ATGGATGTCG AGACAGTCCG
561   CCAAAAACAC TGTCACAAGT AGAGAGGGAC CCTGAGGTCA CACCAGCATC GCGAAGGAGC TCCATGGGCC
631   CCAAATCAAC TTGAGATGAG GCCCGCTTCC CCTGCATTGA CTCCAGAGCC ATCCCCAGTT CCCCATCAAA
701   CACGACCTGA GGCTTGACTC CCTTTAGGCA ACTCCAGAGG TTCCTGGAGA ACAACGGGCC AAGTCTAGAC
771   GAGTACCAAG TTCAGCAGAG CAACTCGAGG ACTGCTCCGT GTACCCCAAA TTGCCTCGAG ATGAGAGCTG
841   ATTCCCTGGC TTAGACTCAA TACGTAGGCC ACCTTTCCAC AAGCCCCTCA AGAGGAGGCT TCTCTCAGCA
911   ATAGGTATGA AGGAGGGACC CTGAGATTGC TGCCTCAATG GGCATATACA CCGAGAAGCC CTGACTCGAA
981   AGGAGGCCGG
```

D. Nucleotide sequence for clone Ov-Msp2:

```
  1   CCGGGTTTCC TTGCAGTGGC TTGAATACAG TCTAGTCTTT CATCTCACAA GGTGAAGGGA TGTCTCACCA
 71   CCTGTGGAGT CCCTAGAGAA AGCCCTAGGT CCCCGACTCC TTTCGACGGA ATGCCTCACA TGCCTCTGAC
141   AACTCGAGAA GCACGCAGAG CTCACTGCTT CAAAATTGGA CGGAGCCTGA CTCCTCTTGA ACATTGACAG
211   GAATCCCAGT ATCCCTGTGG CTACTGGAAA GGGACCCTGG GTCTCCCGAC TCACCTCGAG AAGCGTCCCT
281   TTTGCCCTGC CAAGCCTGCG AAGAGACTCC TGAGGTTGTC CCTCGTAAAC TCGACAGGAC TCCTGACTTC
351   GTTGAACCAA ACACGAGGGT TGGAAGGGCC ATCCCCGCCG TAACTCGAGA ATGTACCCCA GGTCCCCCGC
421   CGCAACTCGA GAAAAACAAT GAGACTTCCC CCTCCAAGCG AGATGAGGTC CAATTCCCCT GCACTGCATG
491   CAGAGCAATT CCGTGAGCCC CCTGAAACAT GTCAGGAGCC TCGACTTGCT TGATGGATGT CGAGACAGTC
561   CGCAAAAACA CTGTCACAAG TAGAGAGGGA CCCTGAGGTC ACATCAGCAT CTCGAAAGAG CTCCATGGGC
631   CCCAAATCAA CTCGAGATGA GGCCCGCTTC CCCTGCATTG ACCCCAGAGC CATCCCCCGT TCCCCATCAA
701   ACACGACCTG AGGCTTGACT CCCTTTAGGC AACTCCAGAG GTTCCCGG
```

E. Nucleotide sequence for clone Aa-Msp1:

```
  1   CCGGTAGGCA ACTCCAGACA TTCCCCGAGN AACACCTTCC CAAGTCTACA GGAACACCAA ATTCAGGACA
 71   GCAAGTGAAC AAAAGCTCCG TGACCCCAAT CCTCTCGAGA TGAAAGCTGA TTTCCTGGCT TCGACTCAGG
141   AGGAGTGCCA ACTTTCCACA AGCACCTCAA GAGGAGGCTT CTTTCATCTA CAGGTATGTC AGACGGACCC
211   TGTTGGTACT GCCTCAAGTG GAATGGACAC CGACATGCTC TCACTCGACA TGAGTCTGGA GTCCCCGGCA
281   GTGACTAGAA TGAAGGCTCG TCTTTCATCT CACCAGATGA GGGGATATCT GAGTCCCCTG TGGAGACCTT
351   ACAGAGAAGC CGTAGTTCCC CGCCTCATCT TGACAGGAGG CCTCCCATCC CTTTGAGAAC TCAAGAGGAT
421   CGNGGAGTGC AATGCTTCCA AAAGAGACGA AGGGGACTCC TAGTGAAAAT GGATAAGAAT CAGAATATCA
491   CTGTGGTAAC AGGAAAGAAG CCCTGAGTCT CGCGCTTCAC CTCGTGAGGC GTCCCTATTG CCCTCGAAAG
561   CCTCCAGGAG AATCCCGAGG TGTCCCTCGC AACGAGACAG GAGTCCTGAT GTCACTGATC CAAACACGAG
631   TTTTGAAGGG ACATCCCCGC CGTAAGTCGA GAATTTACCC CAGGTTCCCG CCGTAACTCG AGAAACACCA
701   TTTAGACTCC CCCCTGGCCG TGAAATGAGG CGCGATTCCC CTGCATAGCG TGCAGAGCAA TTCCGAGTTC
771   CACATCCAAC ACGAAAGGAC CCTTGACTTC CTTGATGGAA CTACAGAGAG TCCCCGAGAA CATTGTCCAA
841   AGTCCCCAGG GACACTGACG TCACTGCAGC CACCCAAAAG TCTTCCCTGT ACGATAAATC AACTCGAGAT
911   GAGGCCAGAT TCCACTGCAC TGGCTCCAGA GCTATTCTGC CTTTCGCATT ACACATAGCT AGTGGCTTCA
981   CGTCCGG
```

F. Nucleotide sequence for clone Aa-Msp2:

```
  1   CCGGCAGTGA CTAGAATGAA GGCTCGTCGT TCATCTCACC AGATGAGGGG ATATCTGAGT CCCCTGTGGA
 71   GACCTTACAG AGNAACCGTA GTTCCCAGCC TCATCTCGAC AGGAGGCCTC CCATCCCTTT GCGAACTCAA
141   GAGGNTCGCG GAGTGCAATG CTTCCAAAAG AGACGAAGGG GACTCCTAGT GAAAATGGAT AAAAATCAGA
211   ATATCACTGT GGGAACAGGA AAGAGGCCCT GGGTCTCGCG CCTCACCTCG TGAGACGTCC CTATTGCCCT
281   CCAAAGCCTC CAGGAGAATC CCGAGGTGTC CCTCGCAACG AGACAGGAGT CCTGATGTCA CTGATCAAAC
351   ACGAGTTTTG AAGGGACATC CCCGTCGTAA GTCGAGAATT TACCCCAGGT TCCCGCCGTA ACTCGAGAAA
421   CACCACTTAG ACTCCCCCCT GGCCGTGAGA TGAGGCGCGA TTCCCCTGCA TAGCGTGCAG AGCAATTCCG
491   AGTTCCACAT CCAACACGAG AGGACCCTTG ACTTCCTTGA TGGAACTACA GAGAGTCCCC GAGAACATTG
561   TCCAAAGTCC CCAGGGACAC TGACGTCACT GCAGCCACCC AAAAGTCTTC CCTGTACGAT AAATCAACTC
631   GAGATGAGGC CAGATTCCAC TGCACTGGCT CCAGAGCTAT TCTGCCTTTT GCATTCTACA TGGCTAGTGG
701   CTTCACGTCC GG
```

G. Nucleotide sequence for clone Oh-Msp1:

```
  1   CCGGCTTTCC TTGCAGTGGC TTGAATACAG TCTAGTCTTT CATCTCACAA GCTGAAGGGA TGTCTGCACC
 71   ACCTGTGGAG TCCCTAGAGA AAGCCCTAGG TCCCCGACTC CTTTCGACGG AATGCCTCAC ATGCCTCTGA
141   CACCTCGAGA GGCACGCAGA GTTCACTGCT TCAAAATGTG ACGGAGCCTG ACTCCTCTTG AACATTGACA
211   GGAATCCCAG TATCCCTGTG GCTACTGGAA AGGGACCCTG GGTTTCCCGA CTCACCTCGA GAAGCGTTCC
281   TTTTGNCCTG CCAAGCCGCG AAGAGACTCC TGAGNTGTCC CTCGTAACTC GACAGGAGTC CTGACTTCGT
351   TGNACCAACA CGAGGGTTGG AAGGGNCATT CCCGCCGTAA GTCGAGAATN NNCCCCAGGT CCCTGCCGCA
421   ACTCGAGAAA AACCATGAGA CTTCCCCCTC CAAGCGAGAT GAGACCCGAT TCCCCTGCAC TGCATGCAGA
491   GCAATTCCGT TAGCCCCCTG AAACATGTCA GGAGCCTCGA CTTGCTTGAT GGGTGTCGAG ACAGTCCGCA
561   AAAACACTCT CACGAGTAGA GAGGGACCCT GAGGTCACAC CAGCATCGCG AAGGAGCTCC ATGGCCCCCA
631   AATCAACTCG AAGATGAGGC CCGCTTCCCC TGCATTTGAC TCCAGAGCTA ATCCCCCGTT CCCCATCCAA
701   ACACGACCTG AGGCTTGACT CCCTTTAGGC AACTCCAGAC GTTCCCCGAG AACAACGGGC CAAGTCTAGA
771   GGAGTACCAA GTTCAGCAGA GCAACTCGAG GACTGCTCCG TGTACCCCAA ATTGCCTCGA GATGAGAGCT
841   GATTCCCTGG CTTAGACTCA AGACGTAGGC CACCTTTCCA CAAGCCCCTC AAGAGGAGGC TTCTCTCAGC
911   AATAGGTATG AGGGAGGGAC CCTGAGATTG CTGCCTCAAT GGGCATATAC ACCGAGAAGC CCTGACTCGA
981   AAGGAGGCCG G
```

H. Nucleotide sequence for clone Oh-Msp2:

```
  1   CCGGCTTTCC TTGCAGTGGC TTGAATACAG TCTAGTCTTT CATCTCACAA GCTGAAGGGA TGTCTGCACC
 71   ACCTGTGGAG TCCCTAGAGA AAGCCCTAGG TCCCCGACTC CTTTCGACGG AATGCCTCAC ATGCCTCTGA
141   CACCTCGAGA GGCACGCAGA GCTCACTGCT TCAAAATTTG ACGGAGCCTG ACTCCTCTTG AACATTGACA
211   GGAATCCCAG TATCCCTGTG GCTACTGGAA AGGGACCCTG GGTCTCCCGA CTCACCTCGA GAAGCGTTCC
281   TTTTGCCCTG CCAAGCCTCG AAGAGACTCC TGAGGTGTCC CTCGTAACTC GACAGGAGTC CTGACTTCGT
351   TGAACCAACA CGAGGGTTGA AGGGCCATCC CCGCCGTAAG TCGAGAATGT ACCCCAGGTC CCCGCCGCAA
421   CTCGAGAAAA ACCATGAGAC TTCCCCCTCC AAGCGAGATG AGGCCCGATT CCCCTGCACT GCATGCAGAG
```

H. Nucleotide sequence for clone Oh-Msp2 (continued):

```
 491   CAATTCCGTG AGCCCCCTGA AACATGTCAG GAGCCTCGAC TTGCTTGATG GGTGTCGAGA CAGTCCGCAA
 561   AAACACTGTC ACAAGTAGAG AGGGACCCTG AGGTCACACC AGCATCGCGA AGGAGCTCCA TGGGCCCCAA
 631   ATCAACTCGA GATGAGGCCC GCTTCCCCTG CATTGACTCC AGAGCCATCC CCCGTTCCCC ATCAAACACG
 701   ACCTGAGGCT TGACTCCCTT TAGGCAACTC CAGAGGTTCC CCGAGAACAA CGGGCCAAGT CTAGAGGAGT
 771   ACCAAGTTCA GCAGAGCAAC TCGAGGACTG CTCCGTGTAC CCCAAATTGC CTCGAGATGA GAGCTGATTC
 841   CCTGGCTTAG ACTCAAGACG TAGGCCACCT TTCCACAAGC CCCTCAAGAG GAGGCTTCTC TCAGCAATAG
 911   GTATGAGGGA GGGACCCTGA GATTGCTGCC TCAATGGCAT ATAACACCGA GAAGCCCTGA CTCGAAAGGA
 981   GGCCGGCTTT CCTTGCAGTG GCTTGAATAC AGTCTAGTCT TTCATCTCAC AAGCTGAAGG GATGTCTGCA
1051   CCACCTGTGG AGTCCCTAGA GAAAGCCCTA GGTCCCCGAC TCCTTTCGAC GGAATGCCTC ACATGCCTCT
1121   GACACCTCGA GAGGCACGCA GAGCTCACTG CTTCAAAATT TGACGGAGCC TGACTCCTCT TGAACATTGA
1191   CAGGAATCCC AGTATCCCTG TGGCTACTGG AAAGGGACCT TGGGTCTCCC GACTCACCTC GAGAAGCGTT
1261   CCTTTTGCCC TGCCAAGCCT CGAAGAGACT CCTGAGGTGT CCCTCGTAAC TCGACAGGAG TCCTGACTTC
1331   GTTGAACCAA CACGAGGGTT GAAGGGCCAT CCCCGCCGTA AGTCGAGAAT GTACCCCAGG TCCCCGCCGC
1401   AACTCGAGAA AAACCATGAC ACTTCCCCCT CCAAGCGAGA TGAGGCCCGA TTCCCCTGCA CTGCATGCAG
1471   AGCAATTCCG TGAGCCCCCT GAAACATGTC AGGAGCCTCG ACTTGCTTGA TGGGTGTCGA GACAGTCCGC
1541   AAAAACACTG TCACAAGTAG AGAGGGACCC TGAGGTCACA CCAGCATCGC GAAGGAGCTC CATGGGCCCC
1611   AAATCAACTC GAGATGAGGC CCGCTTCCCC TGCATTGACT CCAGAGCCAT CCCCCGTTCC CCATCAAACA
1671   CGACCTGAGG CTTGACTCCC TTTAGGCAAC TCCAGAGGTT CCCCGAGAAC AACGGGCCAA GTCTAGATCT
1751   AGAGGAGTAC CAAGTTCAGC AGAGCAACTC GAGGACTGCT CCGTGTACCC CAAATTGCCT CGAGATGAGA
1821   GCTGATTCCC TGGCTTAGAC TCAAGACGTA GGCCACCTTT CCACAAGCCC CTCAAGAGGA GGCTTCTCTC
1891   AGCAATAGGT ATGAGGGAGG GACCCTGAGA TTGCTGCCTC AATGGGCATA TACACCGAGA AGCCCTGACT
1961   CGAAAGGAGG
```

I. Nucleotide sequence for clone Oh-Msp3:

```
   1   CCGGGTTTCC TTGCAGTGGC TTGAATACAG TCTAGTCTTT CATCTCACAA GGTGAAGGGA TGTCTGCACC
  71   ACCTGTGGAG TCCCTAGAGA AAGGCCTAGG TCCCCGACTC CTTTCGACGG AATGCCTCAC ATGCCTCTGA
 141   CAACCTCGAG AAGCAACGCA GAGCTCACTG CTTCAAAGTT GACGGAGCCT GACTCCTCTT GAACATTGAC
 211   AGGAATCCCA GTATCCCTGT GGCTACTGGA AAGGGACCCT GGGTCTCCCG ACTCACCTCG AGAAGCGTCC
 281   CTTTTGCCCT GCCAAGCCTC GAAGAGACTC CTGAGGTGTC CCTCGTAACT CGACAGGACT CCTGACTTCG
 351   TTGAACAAAC ACCAGGGTTG AAGGGCCATC CCCACCGTAA GTCGAGAATG TACCCCAGGT CTCCGCGCAA
 421   CTCGAGAAAA ACCATGAGAC TTCCCCCTCC AAGCGAGATG AGGCCCGATT CCCCTGCACT GCATGCAGAG
 491   CAATTCCGTG AGCCCCCTGA AACATGTCAG GAGCCTCGAC TTGCTTGATG GATGTCGAGA CAGTCCGCAA
 561   AAACACTGTC ACAAGTAGAG AGGGACCCTG AGGTCACACC AGCATCGCGA AGGAGCTCCA TGGGCCCCAA
 631   ATCAATTCGA GATGAGGCCC GCTTCCCCTG CATTGACTCC AGAGACATCC CCAGTTCCCC ATCAAACACG
 701   ACCTGAGGCT TGACTCCCTT TAGGCAACTC CAGAGGTTCC CGGAGAACAA CGGGCCAAGT CTAGACGAGT
 771   ACCAAGTTCA GCAGAGCAAC TCGAGGACTG CTCCGTGTAC CACAAATTGC CTCGAGATGA GAGCTGATTC
 841   CCTGGCTTAG ACTCAAGACG TAGGCCACCT TTCCACAAGC CCCTCAAGAG GAGGCTTCTC TCAGCAATAG
 911   GTATGAGGGA GGGACCCTGA GATTGCTGCC TCAATGGGCA TATACACCGA GAAGCCCTGA CTCGAAAGGA
 981   GGCCGGGTTT CCTTGCAGTG GCTTGAATAC AGTCTAGTCT TTCATCTCAC AAGGTTGAAG GGATGTCTGC
1051   ACCACCTGTG GAGTCCCTAG AGAAAGCCCT AGTTCCCCGA CTCCTTTCGA CGGAATGCCT CACATGCTTC
1121   TGCCAACCCN GAGAGGAACG CAGAGCTCAC TGCTTCAAAA TTTGACGGAG CCTGACTCCT CTTGAGCATT
1191   GTCAGGAATC CCAGTATCCC TGTGGCTAAC TGGAAAGGGA CCTGGGTCTC CCGATTCAAC CTCGAGAAGC
1261   GTCCCTTTTG CCCTGCCAAG CCTCGAAGAG ACTCCTGAGG TGTCCCTCGT AACTCGACAG GAGTCCTGAC
1331   TTCGTTGAAC AAACACGAGG GTTGAAGGGC CATCCCCGCC GTAAGTCGAG AATGTACCCC AGGTCCCCGC
1401   CGCAACTCGA GAAAAACCAT GAGACTTCCC CCTCCAAGCG AGATGAAGCC CGATTCCCCT GCACTGCATG
1471   CAGAGCAATT CCGTGAGCCC CCTGGAAAAC ATGTCAGGAG CCTCGAATTG CTTGATGGAT GTCGAGACAG
1541   TCCGCAAAAA CACTGTCAAC AAGTAGAGAG GGACCCTGAG GTCAACACCA GCATCGCGAA GGAGCTCCAT
1611   GGGCCCCAAA TCAACTCGAG ATGAGGCCCG CTTCCCCTGC ATTGACTCCA GAGCCATCCC CCGTTCCCCA
1671   TCAAACACGA CCTGAGGCTT GACTCCCTTT AGGCAACTCC AGAGGTTCCC CGAGAACAAC GGGCCAAGTC
1751   TAGAGGAGTA CCAAGTTCAG CAGAGAACAC GAGGACTGCT CCGTGTAACC CCAAATTGCC TCGAGATGAG
1821   AGCTGATTCC CTGGCTTAGA CTCAAGACGT AGGCCACCTT TCCACAAGCC CCTCAAGAGG AGGCTTCTCT
1891   CAGCAATAGG TATGAGGGAG GGACCCTGAG ATTGCTGCCT CAATGGGCAT ATACACCGAG AAGCCCTGAC
1961   TCGAAAGGAG GCCGG
```

Fluorescence *in situ* hybridization experiments of each newly isolated satellite DNA clone consistently resulted in pronounced hybridization signals at the centromeric region of all red deer acrocentric chromosomes. These chromosomal deposition patterns were similar to those previously observed with the red deer centromeric satellite DNA clone, Ce-Pst1 (Lee and Lin 1996).

DNA sequence analyses of cervid satellite I DNA.

31 bp subrepeats in monomer clones. Line graphs were produced with data from nucleotide shift self comparisons for all newly isolated representative monomers. Each graph exhibited "in-frame" peaks at approximately every 31 single base shifts, similar to those observed in a previous study (Lee and Lin 1996). This indicated the presence of 31 bp subrepeats in all monomers investigated. A consensus sequence was derived from the subrepeats in each representative monomer clone, each exhibiting a high sequence similarity to the bovine 1.715 satellite subrepeat consensus sequence (Figure 5.2).

The sequence divergence of 31 bp subrepeats among monomers of a given deer species was examined by comparing subrepeats in the mule deer clone Oh-Msp1 with subrepeats in Oh-Msp2a from another mule deer. The results of all possible pairwise sequence comparisons are shown in Table 5.2. Starting with subrepeat 9 of Oh-Msp2a and subrepeat 2 of Oh-Msp1, all consecutive subrepeats in the Oh-Msp2a clone shared identical or near identical sequence identity with subsequent corresponding subrepeats in the Oh-Msp1 clone (Table 5.2, a). Excluding the values from comparisons between these corresponding subrepeats, the remaining pairwise comparisons revealed sequence similarities ranging from 38.7 % to 74.2 % with an average sequence identity of 53.4%. Certain juxtaposed subrepeats in Oh-Msp1, which shared approximately 70% sequence similarity to non-corresponding adjacent subrepeats in Oh-Msp2a, are indicated in Table 5.2 (b - e). Figure 5.3 schematically illustrates those non-corresponding juxtaposed subrepeats which share high sequence similarity. Segment b - c (involving nucleotides

Figure 5.2. Consensus sequences from 31 bp subrepeats of representative cervid monomer clones.

Due to the high sequence similarity of Oh-Msp1 to Ov-Msp1 and Ce-Pst1 to Ce-Msp1, derivation of consensus sequences for representative monomer clones Ov-Msp1 and Ce-Msp1 are not shown. Consensus sequences were produced from strategically aligned subrepeats within clones (**A**) Dd-Pst1, (**B**) Aa-Msp1, and (**C**) Oh-Msp1.

A

```
                                                        Subtracted
Consensus:TCCCTGCCTC AACTCGAGAG GAATCCTGAG T            Nucleotides

   1   ↓↓.AG..TA.↓ ..TG.---.. .---.TC.T. .              C,G,A
  29   .G.A.CT.C.  ..GA...-.. .G..G↓...A .              T
  60   C.....TGGA  G..CAC.... A..GA.CT.. .
  91   ..T.C.....  .T.G..CCC. ..GG...C.C A
 122   .TTT↓AA--   .......... .T-..GC↓.. A              T,G
 152   ...A...T..  C.AAG.G..C ..TG.....C .
 183   --...-.T.G  ..A↓G..T.. ..G...CAG. A              A
 212   ......TGG.  .CG.G..A.. .G.C....G. .
 243   ↓...A-....  .TG.G....C .CG...CA.T .              C
 274   G..A....AA  G.......GA ......C... G↓             TG
 307   .....--.G.  .C..A.GC.. ..G.....↓  .              C
 337   -.G...AACA  ...A..T.T↓ ..↓G..C--- -              CT,GG
 367   ....C...GT  ....T....A T.TA..CC.. G
 398   .T.GG...G.  .........A A..C.A.... A
 429   ↓T..--....  G..AT....T ..GG.....T .              CT
 460   C......↓.G  TG-.GT.... C...T...T. C              A
 491   ..↓AT..AA   .G..-..A.. ..GC.T...T .              A
 523   ...T..ATGG  C....C.... A..A...A.. A
 553   ..AG..T...  .AA..TT... .G.....T.T G
 584   ..A...TAG.  .T.A...A.. .GC...GTG. A
 615   C...AAA-..  .........T ..GAG..T.. .
 645   -....↓TT..  G....A.... ....A.GA.C .              A
 676   .T..ACAAG.  .C..GA.... ..GG.T.CTC .
 707   CAG..A.AG↓  .TG.-..... .G.C...... .              G
 738   .TGTG.....  ..G.G..AT. ..--↓C...  A              A
 768   .G↓..----   -......-.. T..GG.A↓.T A              T,G
 795   ↓......AG                                        N
```

B

```
                                                    Subtracted
Consensus:TCCCTGCCTC AACTCGAGAG GAACCCCGAG T       Nucleotides

   1      C.GG.↓G--. .....C...C ATT....... N        A
  31      A↓↓...T..-. ..G..T.C.. ....A..A.A .        A,A
  63      ..AGGA.AG. ..G.GA.C.A A.G.T...T. A
  94      C...A--A.. CT.......T ...AG.T..T .
 123      ...TG..T.. G....AG... ..GTG..A.C .
 155      .T..ACAAG. .C...A.... ..GG.T↓TTT C        T
 186      ↓.T↓A.G-.- .TG..-...C .G....T.TT G        A,A
 216      GTA....... ..G.G..AT. ..-.A....C A
 246      .G.-.--... .-.....C.T ..GT.TG... .
 273      C...G..AGT G...A..-.T ...G↓T..TC .        G
 304      .T.A.--... .C.↓-...G. ..TAT.T... .        AGA
 335      C.....TGGA G..↓AC.... -..G..GT.. .        CT
 367      ....C..... .T..T..C.. ..G↓.↓.C.- .        G,T
 399      -....TTGAG .....A.... ..T.GNG... .
 429      G.AA...T.. C.AAA....C ...GGG-..C .
 459      --...AGTGA ..A.G..T.A ...T.A-..A .
 487      ↓.A...TGGT ...AG..A.. A.G...T... .        A
 519      ↓.G.-..T.. .C....T... .CGT...T.T .        C
 550      G....--.GA ..↓..C..GA ...T...... G        GC
 581      .↓....--.G. ...GA..C.. ..GT..T..T G        GT
 612      ..A...↓.A- ...A....TT↓ ...GG--..C A        AT,TT
 644      ....C...GT ..G......A↓ TT.....AG. .        T
 676      ....-...GT .........- -...A..ATT .↓       AGAC
 708      ....C-...G G..GT..A.T ..GG.G...T .
 738      C......A.- .G.GT.↓... C..TT..... .        C
 769      ...A--.A.↓ ...A...A.. ..C..TT..C .        C
 799      ...T..ATGG ....AC.... AGT....... A
 830      A.AT..T.CA ..G..CCC.. .G..A.T..C G
 861      ..A....AG. C..C.A.A.. TCT↓..T.↓↓ A        T,T,C
 895      .AAA.----. .........T ..GG..A..T .
 922      C.A....↓.G G-...C.... CT..T.T.CC .        A
 953      ↓↓.A.TA.A. .TAG.T..T. .CT..A..TC C↓       T,G,GG
```

C

```
                                                          Subtracted
Consensus:TCCCTGCCTC AACTCGAGAG GAATCCCGAG T      Nucleotides

  1       ---------- ---------- ----..G.CT .
  8       ...T...AGT GG..T..ATA C.G..TA.TC .
 37       .T.A.CT.A. ..↓.-..-.. .G..GT.--- .       G
 66       G.A.CA...G↓ .G↓C.T.... A..G...T.. G      TGG,T
101       ....C.A... CTT....C-. ...↓.↓.ACA .       G,T
133       G.↓..A.A. --........ .C.CG.A... .        T
163       ..A....T.. ..AAT.T..C .G.G..T..C .
194       --...-.T.G ..↓.T..C.. ......↓.TA .       A,A
224       -.....TGG. T...G..A.. .G.C..T.G. .
254       .T..C.A... .C.......A .CG.T..TTT .
285       GN......AA GC.G...-.↓ ..C...T... N↓      A,TG
318       .....-CG.- .......C.. ..G...T..C .
347       ..GT..NAC. ...A....G.↓ ...GGGNC.- .      TTG
380       ....-...GT ..G......A TNNN...C.. G
410       ........G. .........A A..C.AT... A↓      CT
443       ....--.... C.AG.....T ..GA.....T .
472       C.....↓..G C.TG.-.... C...T...TT A       A
503       G...--...G ..↓AT.TC.. ..GC.T...C .       A
533       .G.T..ATGG GTG......C -.G...↓A.A A       G
563       A.A..CTCA. G.G.A..... .G.C..T... G
595       ..A.AC.AG. .T.G...-.. ..G↓..ATG. C       C
626       C...AAA-.. .......↓.T ..GG....CT .       A
657       C......A.T T↓...C.... C↓.....CC. .       G,T
690       ....CAT.CA ...A...CCT ..GG.TT..C .
721       C..T.TAGG. .....C...C .TTC...... A
752       A.AAC.GGC. ..G..T.... ..G.A..A.. .
783       ..A↓.A.AG-. .........-- ..C.↓.↓..T. .    G,G,T
814       A...CAAA.T GC.......T ..GAG.T..T .
845       -.....G..T .↓....A...C .T.GG..ACC .      G
876       .T..↓A...-. --...A.... ..GG.TTCTC .      ACA
907       ---.A..AAT .GG.A↓.↓... .G.C..T... A      T,GG
938       .TG....... ..-.G.GC.T AT.CA..... A
968       A↓....---- -......A.. ..GG..G.           G
```

Table 5.2: Percent sequence divergence of 31 bp subrepeats in 1 kb monomers from mule deer

Oh-Msp1 Subrepeats

Oh-Msp2a Subrepeats		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
1	(26')	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C. c	N. C.	N. C.	N. C.	N. C.	N. C.	N. C. a	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.
2	(27')	N. C.	55.9	54.5	48.6	54.5	59.4	51.5	51.5	54.3	60.6	54.3	54.5	53.1	50.0	55.9	50.0	51.5	51.5	54.5	48.5	**67.6**	60.6	51.4	51.6	60.6	57.6	**100.0**	54.3	50.0	62.9	55.9	54.5	54.5
3	(28')	N. C.	54.5	48.5	51.4	55.6	58.1	55.6	54.5	52.9	58.1	58.3	54.8	39.5	38.9	51.4	66.7	48.4	58.3	41.9	50.0	58.8	**71.9**	56.2	59.4	48.5	52.9	51.5	**100.0**	54.5	59.4	50.0	53.6	62.1
4	(29')	N. C.	54.5	43.8	45.7	54.3	56.2	57.6	50.0	51.5	54.3	54.5	57.1	42.4	51.6	56.2	50.0	41.9	51.4	51.6	45.5	43.8	63.6	**66.7**	52.9	58.8	54.5	50.0	54.5	**100.0**	57.1	51.6	54.5	54.8
5	(30')	N. C.	50.0	63.6	48.6	50.0	53.3	51.5	43.8	45.2	61.3	58.1	59.4	40.0	43.8	48.3	50.0	51.6	51.5	38.7	50.0	62.1	62.5	45.5	60.6	48.3	53.1	62.5	60.0	55.9	**100.0**	48.4	50.0	53.1
6	(31')	N. C.	42.9	47.1	50.0	47.1	51.5	54.1	48.6	62.5	44.1	54.3	41.9	45.5	46.9	50.0	51.6	48.6	43.8	50.0	66.7	54.3	51.4	40.0	44.1	53.1	54.1	55.9	50.0	53.1	50.0	**100.0**	48.4	44.0
7	(32')	N. C.	54.5	40.0	45.9	45.2	48.6	53.1	40.0	50.0	48.5	55.9	52.9	39.4	48.4	66.7	54.8	56.2	48.5	51.4	48.5	45.5	43.8	40.0	51.6	54.3	50.0	46.9	45.7	58.1	45.2	51.4	**93.5**	51.7
8	(1')	N. C. a	51.9	51.9	51.4	51.5	64.5	50.0	53.3	56.2	51.6	50.0	67.7	42.4	50.0	50.0	58.6	54.5	61.3	50.0	53.1	58.1	67.7	47.2	51.6	45.5	45.2	54.5	61.3	50.0	54.3	45.2	45.5	**100.0**
9	(2')	N. C.	**100.0**	53.1	61.3	42.9	52.9	52.6	58.8	50.0	47.1	48.5	54.5	51.4	51.6	57.6	48.5	56.2	45.2	62.5	48.5	54.3	43.8	62.5	48.5	40.6	52.8	55.9	54.5	54.5	50.0	47.1	54.5	46.4
10	(3')	N. C.	53.1	**100.0**	40.6	55.2	55.2	62.5	53.3	53.1	56.1	48.5	57.6	48.1	44.4	46.9	56.7	46.9	46.7	50.0	59.4	57.1	59.3	47.1	59.4	42.9	46.9	58.6	46.9	45.5	60.0	47.1	53.6	51.9
11	(4')	N. C.	61.3	40.6	**100.0**	46.2	50.0	41.7	48.6	56.4	48.7	52.8	45.9	47.4	59.5	52.8	52.8	54.3	56.2	52.6	59.5	51.4	45.7	55.6	42.9	48.6	54.3	48.6	51.4	54.3	51.4	50.0	45.9	50.0
12	(5')	N. C.	54.8	51.4	47.4	**100.0**	58.8	60.0	55.9	45.5	65.6	50.0	62.9	41.2	55.9	57.6	60.6	62.5	56.0	50.0	45.5	60.6	58.8	52.6	58.8	48.6	47.1	55.9	54.3	54.3	53.1	47.1	46.9	51.7
13	(6')	N. C.	50.0	58.6	54.8	57.6	**96.7**	46.9	62.5	60.0	53.1	63.6	51.6	56.2	46.9	63.6	51.6	50.0	58.1	54.5	64.5	53.1	59.4	48.6	47.1	61.3	54.5	56.2	54.8	59.4	54.8	51.5	50.0	68.0
14	(7')	N. C.	52.6	62.5	42.9	62.9	45.7	**96.8**	61.8	51.5	51.6	51.5	66.7	47.0	47.1	51.6	57.6	45.9	61.5	42.9	60.6	58.8	50.0	50.0	63.6	41.9	51.5	52.9	55.6	57.6	54.3	54.1	50.0	53.3
15	(8')	N. C.	58.8	50.0	53.1	61.1	65.6	61.8	**100.0**	59.4	48.5	50.0	70.0	46.9	50.0	50.0	51.6	58.1	68.0	54.5	53.1	46.9	54.5	63.9	48.5	51.5	48.5	51.5	55.9	51.5	45.5	43.8	47.2	55.6
16	(9')	N. C.	52.9	53.1	63.3	47.1	60.0	48.4	56.2	**96.7**	52.9	54.5	62.5	47.1	46.7	61.3	48.4	56.2	43.8	43.8	61.3	48.4	54.5	50.0	45.2	58.1	45.2	52.9	52.9	54.1	45.2	59.4	50.0	61.3
17	(10')	N. C.	45.5	50.0	53.8	62.5	56.2	46.9	50.0	51.6	**96.8**	53.1	57.6	45.7	51.6	55.9	58.1	61.8	62.5	42.4	51.5	58.8	65.6	52.9	54.8	54.3	53.1	59.4	58.1	55.9	57.6	45.5	42.4	51.7
18	(11')	N. C.	51.6	58.6	59.5	54.3	69.7	54.5	52.9	61.8	50.0	**90.3**	62.9	51.5	57.6	67.6	63.6	62.5	53.1	48.4	65.6	62.9	57.6	55.9	59.4	55.9	57.6	57.1	62.9	60.6	57.6	54.3	52.9	59.4
19	(12')	N. C.	54.5	57.6	45.9	62.9	54.8	63.6	61.8	55.9	60.6	57.6	**100.0**	50.0	51.6	60.6	64.5	48.6	64.0	51.5	51.6	62.5	57.6	51.4	58.1	50.0	48.4	54.5	54.8	58.3	59.4	41.9	52.9	71.4
20	(13')	N. C.	50.0	53.1	46.3	48.6	54.8	51.5	48.4	46.9	50.0	47.1	54.8	**90.9**	44.1	56.2	51.5	45.5	41.4	40.6	55.9	53.1	54.5	45.5	57.6	51.6	50.0	54.5	45.2	48.6	44.4	47.2	43.8	45.5
21	(14')	N. C.	51.6	42.4	62.2	57.6	53.1	43.8	53.3	60.0	54.8	57.6	54.5	47.1	**86.7**	67.7	63.6	59.4	50.0	48.4	59.4	59.4	56.7	58.8	48.4	59.4	58.8	52.9	50.0	61.3	47.1	54.5	56.2	56.7
22	(15')	N. C.	51.5	47.1	50.0	64.7	57.6	54.8	48.5	54.8	57.6	57.6	60.6	45.5	61.8	**93.9**	61.8	58.1	54.8	48.4	51.6	61.3 d	55.9	51.4	57.6	54.8	54.8	58.1	55.6	55.6	51.5	47.1	57.6	45.5
23	(16')	N. C.	50.0	56.7	51.4	58.8	54.8	58.1	51.6	50.0	58.1	60.6	64.5	45.5	56.7	54.8	**96.6**	56.2	61.3	45.7	59.4	61.3	**71.0**	52.6	70.4	53.1	58.1	54.5	63.6	52.9	51.6	44.4	53.1	61.5
24	(17')	N. C.	56.2	48.5	51.4	62.5	54.5	45.7	57.6	53.1	55.9	54.5	57.1	47.1	52.9	64.5	53.1	**96.8**	57.1	54.8	50.0	56.2	50.0	**70.6**	51.6	60.0	48.5	54.5	45.2	41.9	51.5	51.4	58.8	57.7
25	(18')	N. C.	45.2	48.4	56.2	62.9	58.1	69.7	62.5	51.6	51.5	51.5	68.8	43.8	60.0	55.9	61.3	60.6	**100.0**	50.0	51.5	54.8	57.1	52.9	63.0	51.5	48.5	51.5	58.3	51.4	51.5	43.8	45.5	58.6
26	(19')	N. C.	57.6	50.0	52.6	51.5	53.1	47.1	54.5	43.8	45.5	48.6	51.4	42.1	45.2	48.5	47.1	54.8	50.0	**100.0**	45.5	41.9	46.9	50.0	40.6	54.8	51.4	54.5	42.4	50.0	41.9	50.0	55.6	51.7
27	(20')	N. C.	42.4	59.4	56.4	42.4	64.5	66.7	53.1	67.7	46.9	63.6	51.6	48.5	54.8	64.5	62.5	53.1	52.9	38.7	**93.5**	52.8	52.9	50.0	58.1	57.6	64.7 e	51.5	50.0	56.2	42.9	67.6	51.6	57.6
28	(21')	N. C.	43.8	58.1	54.3	58.8	50.0	58.8	46.9	51.6	60.6	62.9	59.4	54.8	56.2	61.3 b	61.3	54.8	58.1	41.9	58.8	**96.8**	54.5	47.4	51.5	50.0	55.9	**69.7**	58.8	43.8	59.4	54.3	40.6	54.8
29	(22')	N. C.	47.1	48.4	47.4	60.0	61.3	48.4	56.2	54.8	64.5	50.0	59.4	41.9	46.7	54.5	**70.0**	51.6	57.6	48.4	51.5	56.2	**96.8**	58.8	63.0	58.1	51.6	63.6	**74.2**	64.5	64.5	52.9	43.8	64.3
30	(23')	N. C.	53.1	40.6	56.8	51.4	57.6	51.4	61.8	53.1	53.1	53.1	51.5	39.4	54.8	55.9	52.9	**74.2**	57.1	51.5	50.0	50.0	56.2	**90.9**	48.4	47.1	48.5	52.9	53.1	61.3	41.9	41.2	54.5	51.6
31	(24')	N. C.	45.9	61.3	42.9	56.2	50.0	64.7	50.0	41.9	59.4	48.4	67.7	47.1	50.0	51.6	66.7	51.5	**70.0**	41.7	48.6	57.6	62.5	48.6	**96.3**	50.0	38.7	53.1	61.8	54.3	57.6	45.5	53.1	55.6
32	(25')	N. C.	44.4	46.4	54.8	41.9	64.5	44.1	53.1	64.5	54.5	54.3	52.9	51.6	50.0	61.3	53.1	58.1	51.5	51.6	57.6	50.0	56.2	62.2	55.6	**96.8**	51.6	58.8	48.5	55.9	51.6	58.8	51.4	53.3
33	(26')	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	**100.0**	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.	N. C.

a - Sequence comparisons of corresponding subrepeats (bold boxes).

b, c - Regions of 3 adjacent subrepeats sharing approximately 70% sequence similarity.

d, e - Regions of 2 adjacent subrepeats sharing approximately 70% sequence similarity.

Subrepeats in Oh-Msp2a are renumbered (in brackets) according to their near identical sequences to corresponding Oh-Msp1 subrepeats.

466 - 650 of Oh-Msp1) have approximately 70% sequence similarity to region b' - c' (involving nucleotides 869 - 985 and 1 - 63 of Oh-Msp2a). Since nucleotides 869 - 985 and 1 - 63 of Oh-Msp2a are analogous to nucleotides 651 - 837 of Oh-Msp1, nucleotides 466 - 837 of Oh-Msp1 could constitute a tandem duplication of 186 bp (See top panel of Figure 5.3). Likewise, about 70% sequence similarity is demonstrated between the DNA segments d - e and d' - e', again substantiating the notion that DNA regions corresponding to nucleotides 466 - 837 of Oh-Msp1 could represent a 186 bp tandem duplication.

Intragenomic sequence conservation. Sequence similarities between cervid satellite I monomers within an individual animal were determined in three deer species. Based on the sequence identity of clone Ov-Msp2 to the first 753 bp of Ov-Msp1 from a white tailed deer genome, an intragenomic sequence conservation of 95.6% was calculated. Mismatches between Ov-Msp1 and Ov-Msp2 consisted of 22 single base substitutions as well as 6 single base insertions, 3 single base deletions, and a 2 nucleotide deletion in Ov-Msp2 (Figure 5.4A). Likewise, in a North American moose, the sequence for the last 712 bp of clone Aa-Msp1 showed 97.1% sequence similarity with clone Aa-Msp2. Mismatches during this comparison consisted of 20 single base substitutions and 1 single base insertion in Aa-Msp1 (Figure 5.4B). Clones Oh-Msp2 and Oh-Msp3 provided the complete DNA sequences of four different 1 kb monomers from the same mule deer. All possible pairwise comparisons between these four 1 kb monomers revealed sequence similarities from 96.4 % to 99.6 % with an average of 97.6 % (Table 5.3). Approximately, one half of the mismatches observed during these sequence comparisons consisted of single base substitutions and the other half of the mismatches were single base insertions / deletions (Data not shown).

Intraspecific sequence conservation. DNA sequence similarities of over 95% were also consistently achieved during intraspecific sequence comparisons between the Oh-Msp1

Figure 5.3. A 0.18 kb tandem duplication in 1 kb monomers of Oh-Msp1 and Oh-Msp2a having a higher-order organization of 31 bp subrepeats.

Subrepeats 16 - 18 and 21 - 23 of Oh-Msp1 share approximately 70% sequence similarity with subrepeats 22' - 24' and 27' - 29' of Oh-Msp2a and are indicated by speckled bars b, c, b', c', respectively. Subrepeats 22 - 23 and 27 - 28 of Oh-Msp1 also share 70% sequence similarity with subrepeats 16' - 17' and 21' - 22' of Oh-Msp2a and are indicated by solid bars d, e, d', e', respectively. The position of nucleotides which define the 0.18 kb duplication are indicated in the Oh-Msp1 and Oh-Msp2a monomers. The tandem duplication region is also illustrated in the top panel by boxes filled with diagonal lines.

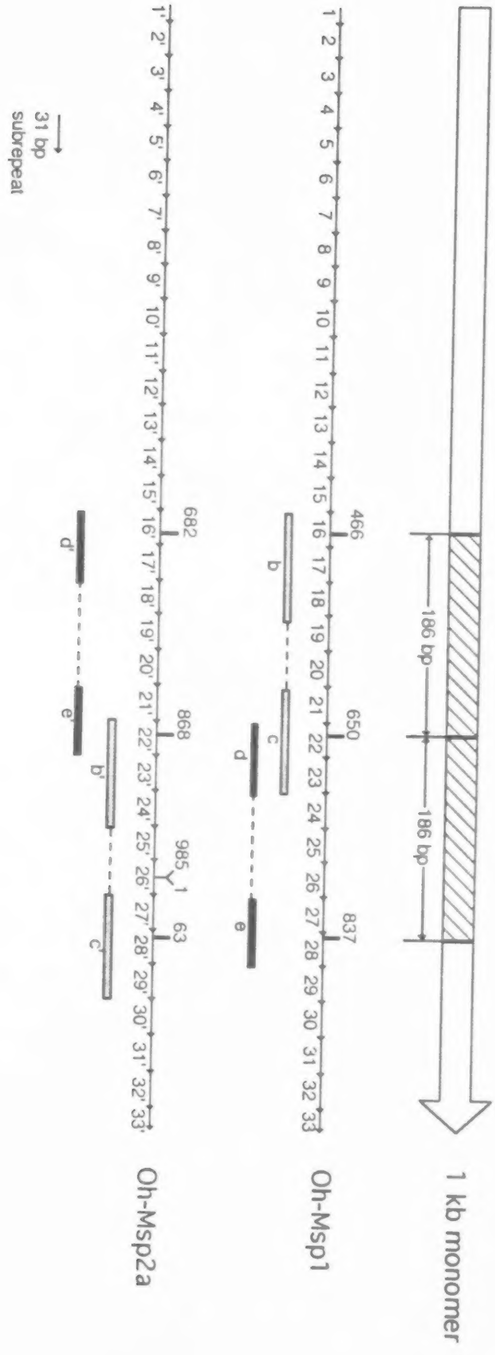

1 kb monomer
186 bp
186 bp
Oh-Msp1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33
466
650
837
b
c
d
e
Oh-Msp2a
1' 2' 3' 4' 5' 6' 7' 8' 9' 10' 11' 12' 13' 14' 15' 16' 17' 18' 19' 20' 21' 22' 23' 24' 25' 26' 27' 28' 29' 30' 31' 32' 33'
682
868
985
1
63
d'
e'
b'
c'
31 bp
subrepeat

Figure 5.4. Intragenomic and intraspecific monomeric sequence comparisons.

(**A**) Intragenomic sequence comparisons between the first 753 bp of Ov-Msp1 with Ov-Msp2 and (**B**) the last 712 bp of Aa-Msp1 with Aa-Msp2. (**C**) Intraspecific sequence comparisons between Oh-Msp1 and Oh-Msp2a. Single base substitutions are denoted either by C, T, A, G and single base insertions are designated by vertical arrows.

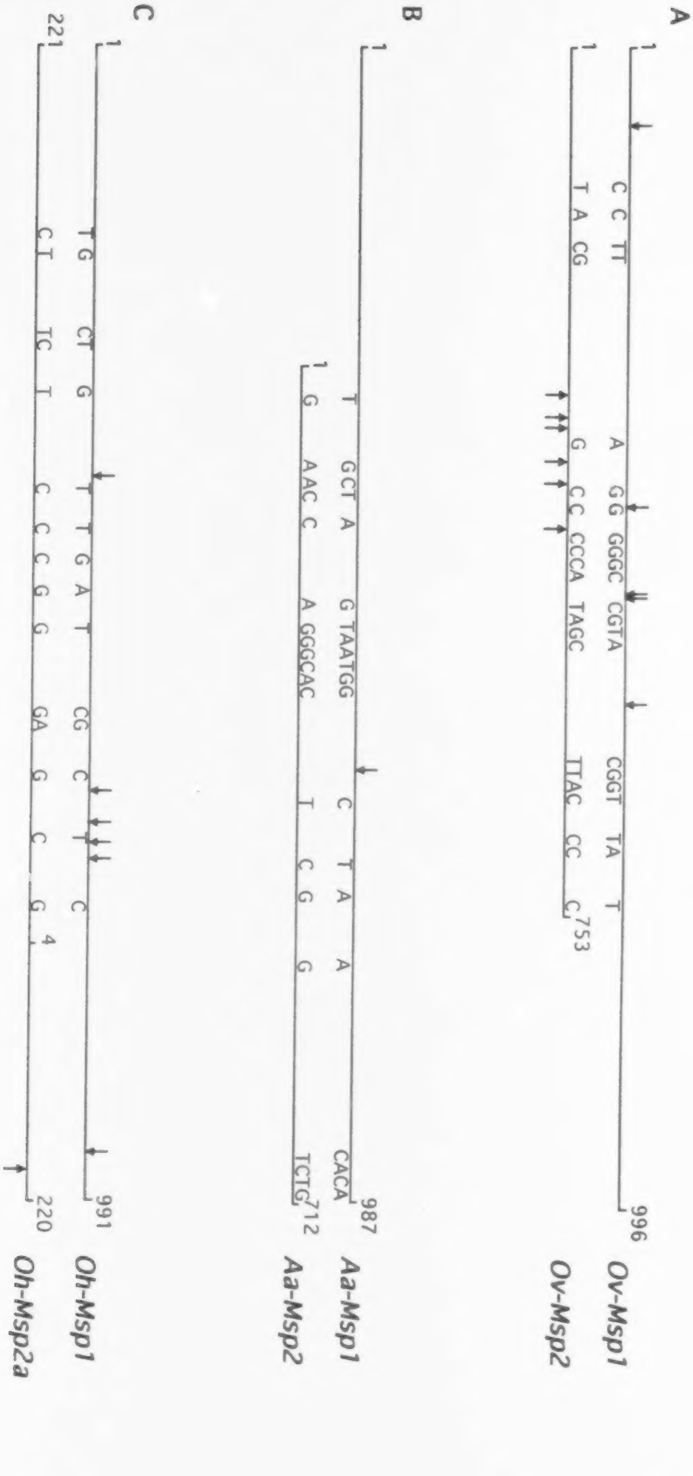
A
B
C
Ov-Msp1
Ov-Msp2
Aa-Msp1
Aa-Msp2
Oh-Msp1
Oh-Msp2a
1
996
753
987
712
991
221
220
4

Table 5.3. Intragenomic and Intraspecific Sequence Similarities

	Clones compared		Sequence Similarity
	Clone 1	Clone2	
Intragenomic:	Ov-Msp1	Ov-Msp2	95.8%
	Aa-Msp1	Aa-Msp2	97.1%
	Oh-Msp2a	Oh-Msp2b	99.6%
	Oh-Msp3a	Oh-Msp3b	96.4%
	Oh-Msp2a	Oh-Msp3a	97.1%
	Oh-Msp2a	Oh-Msp3b	97.3%
	Oh-Msp2b	Oh-Msp3a	97.4%
	Oh-Msp2b	Oh-Msp3b	97.7%
Intraspecific:	Oh-Msp1	Oh-Msp2a	97.4%
	Oh-Msp1	Oh-Msp2b	97.8%
	Oh-Msp1	Oh-Msp3a	97.1%
	Oh-Msp1	Oh-Msp3b	95.6%

clone of a mule deer and individual monomers of the Ov-Msp2 and Ov-Msp3 clones from another animal of the same species (Table 5.3). Approximately three quarters of the mismatches observed during these comparisons consisted of single base substitutions and one quarter consisted of single base insertions / deletions (e.g. Figure 5.4C).

Interspecific sequence conservation. The sequence alignment strategies for interspecific sequence comparisons between the 10 representative monomer clones from 10 different cervid species are shown in Figure 5.5 and the results presented in Table 5.4. Maximum sequence similarity was achieved between the 0.8 kb monomers and 1 kb monomers when a specific region of approximately 186 bp was first removed from each 1 kb monomer. Furthermore, maximum sequence similarities in comparisons with the CCSatI clone of the roe deer, was achieved when the CCSatI clone was treated as a partial sequence of a 1 kb monomer. Specifically, a gap of 260 bases was required after nucleotide 730 and nucleotides 219 - 402 (i.e. a 184 bp region) were removed from this clone before sequence comparisons. Relatively high interspecific sequence similarities were observed between clones Oh-Msp1 (mule deer) and Ov-Msp1 (white tailed deer) (94.7%), Ce-Msp1 (wapiti) and Ce-Pst1 (red deer) (92.3%), Dd-Pst1 (fallow deer) and Ce-Pst1 (red deer) (87.8%), and Dd-Pst1 (fallow deer) and Ce-Msp1 (wapiti) (87.3%).

Discussion:

Presence and organization of 31 bp subrepeats in cervid satellite I monomer clones.

The presence of 31 bp subrepeats in cervid satellite I DNA was initially identified in an Indian muntjac satellite DNA clone (Bogenberger *et al.* 1985). Similar subrepeats have also been observed in a number of other deer species (Lee and Lin 1996). Here, 31 bp subrepeats were also found in newly isolated cervid satellite I monomer clones from five other deer species, suggesting that these 31 bp subrepeats exist in satellite I DNA monomers of all deer species.

Figure 5.5. Alignment strategies for interspecific sequence comparisons of 10 representative cervid centromeric satellite I DNA monomer clones.

Certain nucleotide positions for each monomer are indicated and a 260 bp gap in the CCSatI DNA sequence is represented by a dotted line. Specific 0.18 kb DNA sequences, which were removed from the 1 kb monomers to obtain maximum sequence similarities with 0.8 kb clones, are denoted by triangles.

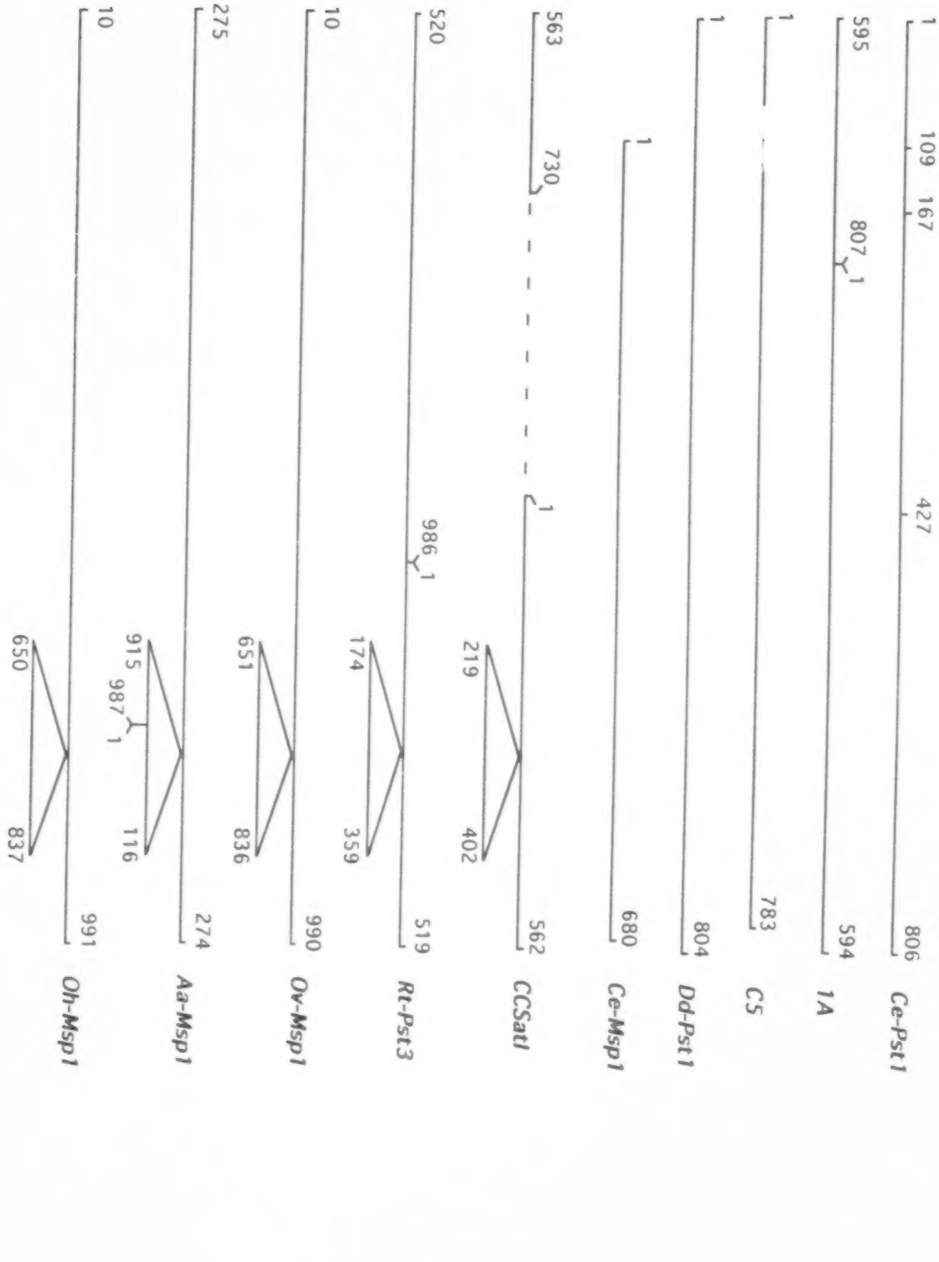

Ce-Pst1
1A
C5
Dd-Pst1
Ce-Msp1
CCSatI
Rt-Pst3
Ov-Msp1
Aa-Msp1
Oh-Msp1

Table 5.4. Interspecific Sequence Homology Among 10 Representative Centromeric Satellite DNA Clones

	1A	C5	Dd-Pst1	Ce-Msp1	CCSatI	Rt-Pst3	Ov-Msp1	Aa-Msp1	Oh-Msp1
Ce-Pst1	77.6	74.7	87.8	92.3	73.0	78.9	78.1	77.6	78.2
1A	---	84.9	71.9	77.9	69.5	75.7	73.9	74.8	74.4
C5		---	75.3	74.7	67.1	73.1	72.0	72.7	72.3
Dd-Pst1			---	87.3	73.2	76.5	77.4	77.2	77.4
Ce-Msp1				---	71.6	77.9	77.3	77.5	77.9
CCSatI					---	74.5	73.1	72.8	73.6
Rt-Pst3						---	84.8	76.3	85.2
Ov-Msp1							---	73.9	94.7
Aa-Msp1								---	74.3

Often the presence of these subrepeats go undetected because of the extent of their nucleotide sequence variation. In the present study, detailed sequence comparisons among 31 bp subrepeats of monomers from two mule deer revealed an average sequence divergence of 46.6%. However, the nth subrepeat of each monomer exhibits near or complete sequence identity suggesting that each cervid monomer represents a hierarchical organization of these subrepeats. Higher-order repeats in mammalian centromeric satellite DNA families usually exhibit much more sequence homogeneity than their constituent basic repeat units (Willard and Waye 1987). Thus, in the context of cervid monomers as higher-order repeats, it is not surprising that extremely high intragenomic and intraspecific sequence similarities were observed between cervid satellite I DNA monomers. Similarly, the 1.4 kb monomers of bovine satellite I likely also represent a higher-order repeat of diverged 31 bp subrepeats since less than 3% sequence divergence was observed between independently isolated monomer clones (Taparowsky and Gerbi 1982).

Genesis of 0.8 kb and 1 kb cervid satellite I DNA monomers.

Cervid satellite I DNA is primarily organized into monomers of 0.8 kb in plesiometacarpalian deer and 1 kb in telemetacarpalian deer. Interspecific sequence comparisons between 0.8 kb and 1 kb monomers demonstrated that the additional 0.18 kb DNA is localized to a specific region within the 1 kb monomers (Figure 5.5). Intraspecific sequence comparisons of subrepeats from different 1 kb monomers revealed that each 1 kb monomer contains two adjacent 0.18 kb segments which share approximately 70% sequence similarity (Figure 5.3). Cumulatively, these data support the postulation that 1 kb cervid satellite I DNA monomers are indeed derived from a 0.18 kb tandem duplication of an original 0.8 kb DNA sequence (Lee and Lin 1996).

It is uncommon to find two or more grossly variant sized monomers in a single satellite DNA family. However, if the 0.8 kb and 1 kb cervid satellite I DNA monomers are considered as higher order organizations of subrepeats, then these different sized higher-order repeats could have been produced from unequal crossing over events (Willard and Waye 1987).

It is believed that a primordial 31 bp DNA sequence was amplified before the divergence of bovids and cervids, some 25 million years ago. Approximately 26 tandemly arranged subrepeats (from an initial amplification event) producing a 0.8 kb higher-order DNA unit in a progenitor deer species. A 0.18 kb duplication then occurred in this DNA sequence in telemetacarpalian deer resulting in a 1 kb monomer. No 0.18 kb duplication is thought to have occurred in plesiometacarpalian deer maintaining a 0.8 kb monomer. Further amplification of these monomers account for their multiple presence and tandem organization in the genomes of present day deer species (Figure 5.6).

Implications of intraspecific sequence conservation.

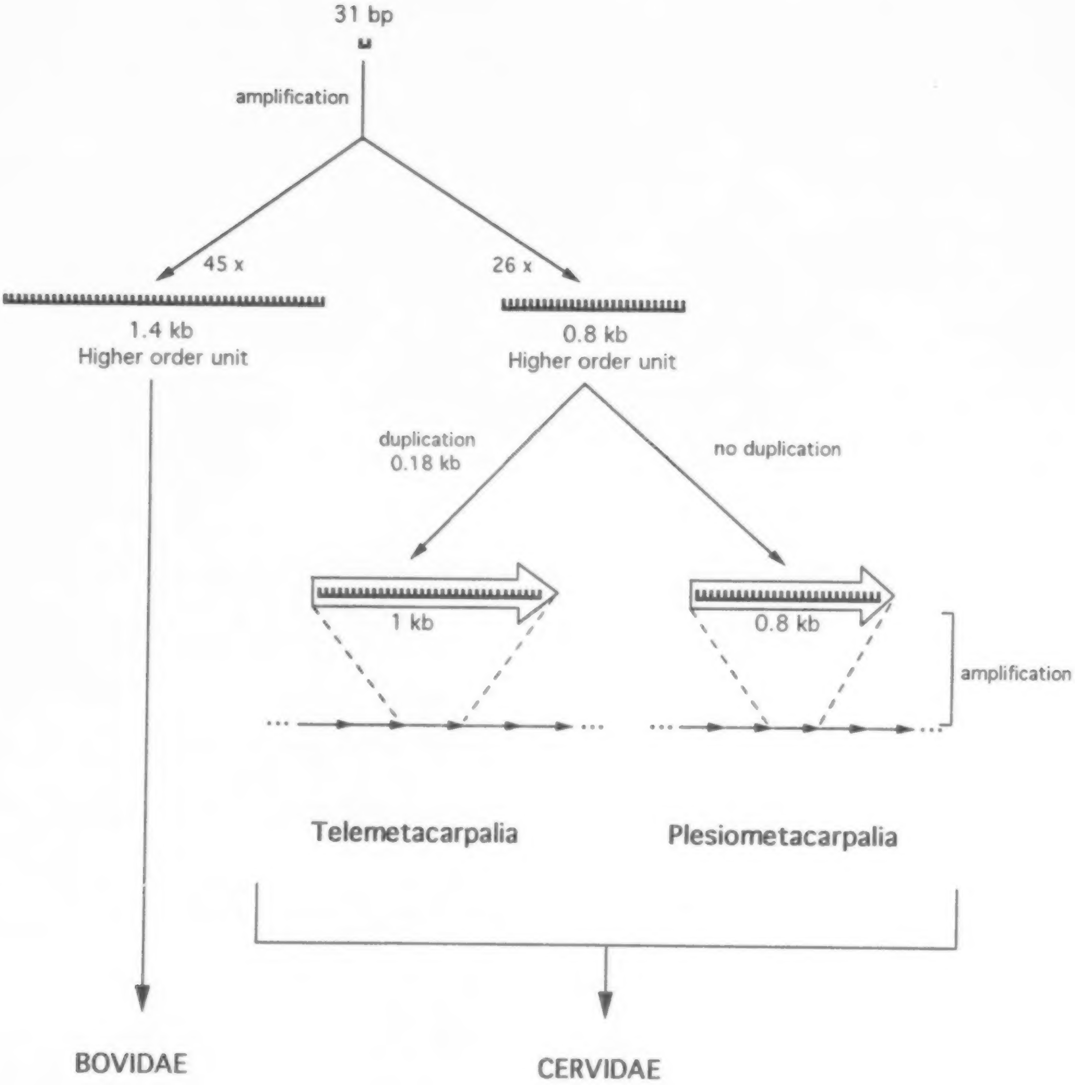
According to Whitehead (1993), there are currently 41 known deer species in the world of which 1 is probably extinct and 6-7 species belong to the genus Muntiacus. Based primarily on distinct geographical distributions rather than specific marked physical differences, 196 subspecies of deer have now been proposed.

Since all cervid satellite I DNA monomers appear to represent higher-order repeat structures, high intraspecific monomer sequence conservation could be observed in deer species (i.e. in mule deer). This validates the use of a single monomer sequence for each deer species during interspecific sequence comparisons for the purposes of establishing phylogenetic relationships.

Results from interspecific sequence comparisons of representative monomers from 10 different deer species imply that the white tailed deer is very closely related to the mule deer. This is consistent with other studies suggesting a close genetic relationship between these two Odocoileus species (e.g. Derr *et al.* 1991). In 1777, Erxleben identified the North American elk (wapiti) as the distinct species, *Cervus*

Figure 5.6. Proposed genesis of satellite I DNA in cervids.

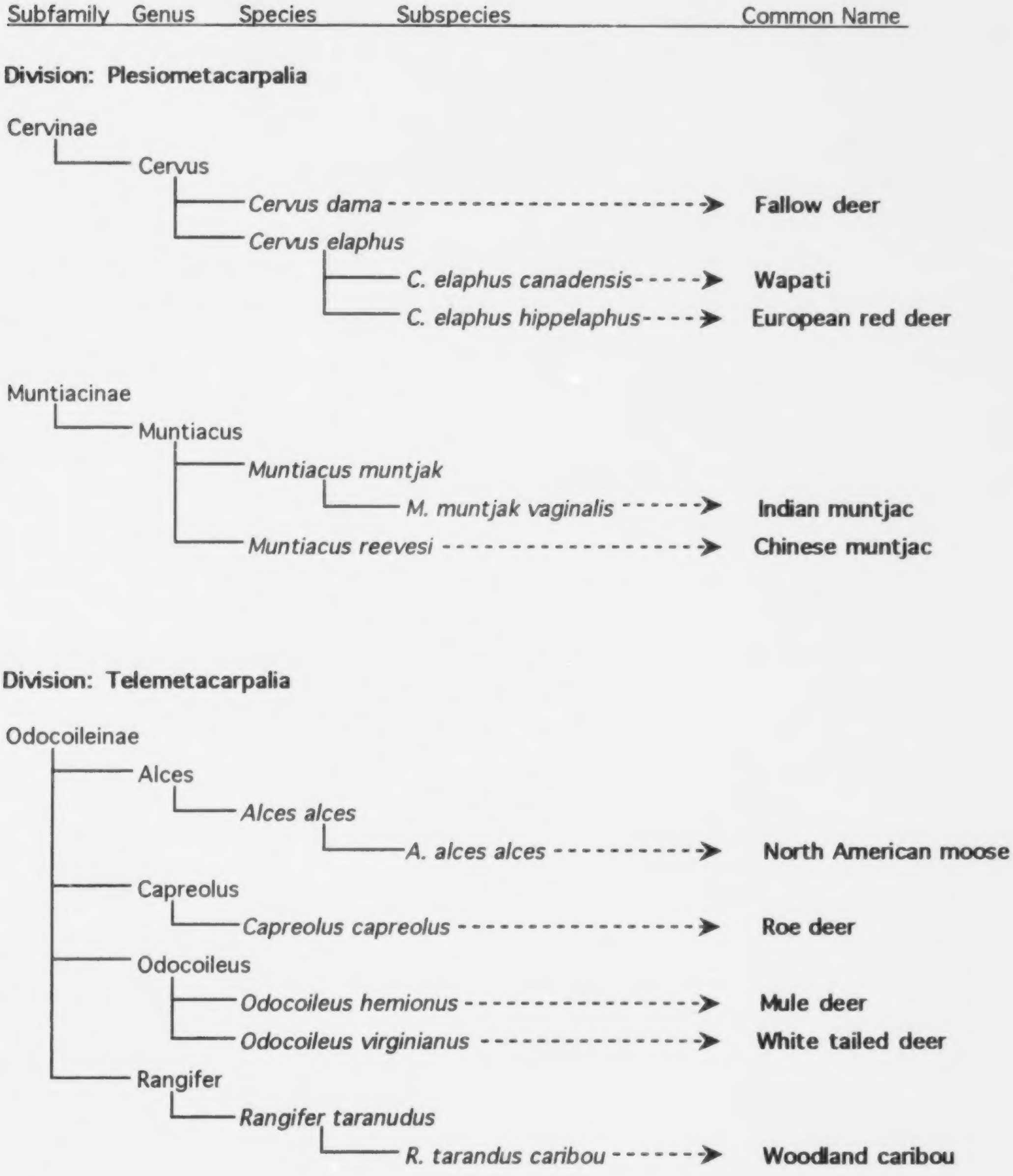
A 31 bp DNA sequence is amplified 26 times to produce a higher-order 0.8 kb DNA unit in an ancestral deer species and 45 times to produce a higher-order 1.4 kb DNA unit in bovids. In telemetacarpalian deer, a 0.18 kb tandem duplication occurs producing a 1 kb monomer. No duplication occurs in plesiometacarpalian deer yielding a 0.8 kb monomer. These monomers are then amplified, accounting for the multiple presence and tandem organization of these two sized monomers in present day deer species.


31 bp
amplification
45 x
26 x
1.4 kb
Higher order unit
0.8 kb
Higher order unit
duplication
0.18 kb
no duplication
1 kb
0.8 kb
amplification
Telemetacarpalia
Plesiometacarpalia
BOVIDAE
CERVIDAE

canadensis (Kurten and Anderson 1980). However, its similar morphological characteristics and ability to readily interbreed with the red deer, have already led others to consider the wapiti as a subspecies of *Cervus elaphus* (Figure 5.7). Furthermore, some authorities argue that the fallow deer should be classified as a species in the genus *Cervus* rather than as a separate genus (e.g. Corbet and Hill 1986). The high degree of sequence similarity between cervid satellite I DNA clones from the red deer, fallow deer, and wapiti, in the present study, suggest a high degree of genetic relatedness between these three deer species and hence substantiates these reclassifications.

Figure 5.7. Taxonomic reclassification of 10 deer species from which centromeric satellite I DNA clones have been isolated.

This classification is based upon the listing found on the NCBI Taxonomy Browser (http://www3.ncbi.nlm.nih.gov/taxonomy/taxonomy_home.html).


Subfamily
Genus
Species
Subspecies
Common Name
Division: Plesiometacarpalia
Cervinae
Cervus
Cervus dama
Fallow deer
Cervus elaphus
C. elaphus canadensis
Wapati
C. elaphus hippelaphus
European red deer
Muntiacinae
Muntiacus
Muntiacus muntjak
M. muntjak vaginalis
Indian muntjac
Muntiacus reevesi
Chinese muntjac
Division: Telemetacarpalia
Odocoileinae
Alces
Alces alces
A. alces alces
North American moose
Capreolus
Capreolus capreolus
Roe deer
Odocoileus
Odocoileus hemionus
Mule deer
Odocoileus virginianus
White tailed deer
Rangifer
Rangifer taranudus
R. tarandus caribou
Woodland caribou

References:

Bogenberger JM, Neumaier PS, Fittler F. (1985) The Muntjak satellite IA sequence is composed of 31-bp-pair internal repeats that are highly homologous to the 31-base-pair subrepeats of the bovine satellite 1.715. *Eur J Biochem* 148: 55-59.

Bogenberger JM, Neitzel H, Fittler F. (1987) A highly repetitive DNA component common to all cervidae: its organization and chromosomal distribution during evolution. *Chromosoma* 95: 154-161.

Brooke V. (1878) On the classification of the Cervidae, with a synopsis of the existing species. *J Zool* 883-928.

Derr JN, Hale DW, Ellsworth DL, Bickman JW. (1991) Fertility in an F_1 male hybrid of white tailed deer (*Odocoileus virginianus*) x mule deer (*O. hemionus*). *J Reprod Fert* 93: 111-117.

Goss RJ. (1983) Deer antlers: regeneration, function, and evolution. New York, Academic Press. p.42-51.

Horz W, Altenburger W. (1981) Nucleotide sequence of mouse satellite DNA. *Nucleic Acids Res* 9: 683-696.

Kurten B, Anderson E. (1980) Pleistocene mammals of North America. New York, Columbia University Press. p.309.

Lee C, Ritchie DBC, Lin CC. (1994) A tandemly repetitive, centromeric DNA sequence from the Canadian woodland caribou *(Rangifer tarandus caribou)*: Its conservation and evolution in several deer species. *Chromosome Res* 2: 293-306.

Lee C, Lin CC. (1996) Conservation of a 31 bp bovine subrepeat in variable sized centromeric satellite DNA monomers of *Cervus elaphus* and other cervid species. *Chromosome Res* 4: 428-436.

Lin CC, Sasi R, Fan Y-S, Chen Z-Q. (1991) New evidence for tandem chromosome fusions in the karyotypic evolution of Asian muntjacs. *Chromosoma* 101: 19-24.

Plucienniczak A, Skowronski J, Jaworski J. (1982) Nucleotide sequence of bovine 1.715 satellite DNA and its relation to other bovine satellite sequences. *J Mol Biol* 158: 293-304.

Rosenberg H, Singer M, Rosenberg M. (1978) Highly reiterated sequences of simiansimiansimiansimiansimian. *Science* 200: 394-402.

Scherthan H. (1991) Characterization of a tandem repetitive sequence cloned from the deer *Capreolus capreolus* and its chromosomal localisation in two muntjac species. *Hereditas* 115: 43-49.

Taparowsky EJ, Gerbi SA. (1982) Sequence analysis of bovine satellite I DNA (1.715 gm/cm^3). *Nucleic Acids Res* 10: 1271-1281.

Willard HF, Waye JS. (1987) Hierarchical order in chromosome-specific human alpha satellite DNA. *Trends Genet* 3: 192-198.

Wilson DE, Reeder DM. (1993) Mammal species of the world. A taxonomic and geographic reference. Washington DC, Smithsonian Press.

Whitehead GK. (1993) The Whitehead Encyclopedia of deer. Shrewsbury, England, Swan Hill Press. p 238-242, 313-4.

CHAPTER 6

CONCLUSIONS

Concluding remarks and implications of this research.

Human gamma satellite DNA was first discovered as a 220 bp tandemly repetitive DNA family with one subfamily being specific for the centromeric region of human chromosome 8 (Lin *et al.* 1993). Copy number estimations suggested that approximately 0.43 Mb of gamma satellite DNA resided in the centromeric region of a single chromosome 8.

Chapter 2 of this thesis expounded on the characterization of this novel human centromeric DNA family, describing the identification and characterization of another gamma satellite subfamily; this subfamily being specific for the X-chromosome. This DNA subfamily was defined as gamma satellite DNA based on the presence of tandemly arranged 220 bp repeat units which shared approximately 62% sequence similarity with gamma 8 satellite monomers. The discovery of this second subfamily of gamma satellite DNA suggests that other subfamilies of gamma satellite DNA may also exist in the human genome. However, the extreme chromosome specificity observed suggests that other putative subfamilies may be undetectable by conventional cross hybridization techniques.

Several features of human gamma satellite DNA were made known from observations described in chapter 2. (1) A contiguous 20 bp region was discovered when derived consensus sequences for gamma 8 and X monomers were compared. The significance of this finding is appreciated in view of the fact that a 17 bp sequence in alpha satellite monomers serves as a binding site for the CENP-B protein (Masumoto *et al.* 1989). (2) At least one inversion is thought to exist in gamma satellite DNA array(s) of the X-chromosome. Three kilobases of sequenced gamma 8 DNA have failed to reveal any inverted monomers as with the gamma X DNAs. (3) Common physical features of both gamma satellite DNA subfamilies (i.e. similar copy number estimations and the lack of RFLPs with two different restriction enzymes) may suggest the imposition of structural and/or functional constraints on this centromeric DNA family. (4) Finally, gamma 8 and X subfamilies are consistently localized as two distinct and opposed fluorescent signals at specific chromosome centromeres. Albeit circumstantial evidence, this may imply that gamma satellite DNA is in close proximity to the kinetochore domain and thus could specifically interact with the kinetochore itself.

To further characterize this new human centromeric satellite DNA family, two further experiments were conducted to ascertain the conservation of gamma satellite DNA in two old world primate species and in a mitotically stable marker chromosome derived from human chromosome 8. In chapter 3, gamma 8 and X satellite DNAs were undeniably shown to be present in the genomes of the African green monkey and the chimpanzee. Southern blot data revealed the repetitive nature of this DNA family in these two primate genomes, however, monomer size could not be established. The presence of gamma satellite DNA in these two old world primates further implies a functional significance for this DNA family.

Fluorescence *in situ* hybridization studies permitted the extrapolation of two main conclusions. First, sufficiently more gamma 8 and gamma X satellite DNA may reside in the genome of the African green monkey than in the genome of humans. The significance of this is unknown, however, similar observations have been previously made for alpha satellite DNA. Second, these DNA sequences remain very chromosome-specific in these two primate species. In fact, it is believed that gamma 8 satellite DNA hybridizes exclusively to the centromeric region of the two primate species' human chromosome 8 counterpart. Likewise the gamma X satellite DNA is thought to be specific for the centromeric region of primate X-chromosomes. Certain authors have suggested that a chromosome-specific nature of centromeric heterochromatin may assist in the recognition of homologous chromosomes during meiotic pairing (e.g. Haaf *et al.* 1986, Choo *et al.* 1988).

Some small stable marker chromosomes were known not to contain detectable alpha satellite DNA sequences. These observations have led to a continuing debate as to how essential alpha satelllite DNA is for the formation of a functional centromere. An attempt was made in chapter 3 of this thesis to examine whether other centromeric satellite DNA sequences (such as gamma 8 satellite DNA) are present in such minute marker chromosomes. The negative results obtained from this study suggest that either other centromeric DNA sequences are responsible for a functional centromere or that small amounts of alpha and/or gamma satellite DNA are actually present but in amounts undetectable by current fluorescent *in situ* hybridization procedures.

Compared to the numerous studies on human centromeric DNAs, very little is known of centromeric DNAs in other mammalian species. In particular, information and understanding of centromeric DNAs in the deer family is extremely limited. Prior to the investigations described in chapters 4 and 5 of this thesis, single monomer clones were obtained, sequenced, and characterized from four different deer species (Bogenberger *et al.* 1985, Yu *et al.* 1986, Lin *et al.* 1991, Scherthan 1991, Lee *et al.* 1994). Southern blot and *in situ* hybridization experiments (e.g. Bogenberger *et al.* 1987, Lee *et al.* 1994) suggested that all these isolated clones belonged to a single major cervid centromeric satellite DNA family (i.e cervid satellite I DNA) which is very well conserved throughout all deer species studied. However, the monomer sizes differed between 0.8 kb in plesiometacarpalian deer and 1 kb in telemetacarpalian deer. Cervid satellite I DNA were even observed as 2 kb monomers in the Siberian roe deer (Scherthan 1991). This raised the question as to how a single repetitive DNA family could have repeat units of such differing sizes.

The key to deciphering this dilemma was in understanding the composition and evolution of this repetitive DNA family. Chapter 4 of this thesis established the presence of tandemly repetitive 31 bp subrepeats in all currently identified cervid satellite I monomers. Although a seemingly trivial observation, this establishes a common origin for the genesis of all cervid satellite I DNA monomers, regardless of repeat unit size.

The relationship of 0.8 kb and 1 kb cervid satellite I DNA monomers was not established prior to 1994. Lee *et al.* (1994) then isolated and sequenced the first 1 kb cervid satellite I DNA monomer from the Canadian woodland caribou and observed a 60% sequence similarity between the first and last 191 bp of the clone. From this data, two speculations were put forward. First, it was suggested that the 1 kb monomers arose from ancestral 0.8 kb monomers, rather than de novo. Second, the 1 kb monomers were formed from selective amplification of a 0.8 kb monomer along with 191 bp of an adjoining 0.8 kb monomer. This second speculation has now been refuted by the results of detailed pair-wise sequence comparisons between subrepeats of different 1 kb cervid satellite I DNA monomers (Table 5.2). A tandem duplication observed in 1 kb cervid satellite I DNA monomers suggests that these monomers were likely formed by a 0.18 kb DNA duplication event in an ancestral 0.8 kb monomer (Chapter 5 and Figure 5.6).

The concept of the 0.8 kb and 1 kb monomers as higher order repeats of 31 bp sequences in cervid satellite DNA provides an explanation for several previously conceived inconsistencies. Studies of mammalian centromeric satellite DNA families have indicated that in a given DNA family, the size of fundamental repeat units is generally consistent. However, these fundamental repeat units can be organized into higher-order repeat units, which can vary in length between certain non-homologous chromosomes, but exhibit extreme sequence homogeneity within a chromosome subfamily (Willard and Waye 1987). Therefore, the presence of two different sized higher-order repeats in cervid species would not be contradictory to current dogma for satellite DNA families and would account for the high intragenomic and intraspecific sequence similarity observed between independently isolated 0.8 kb and 1 kb monomers (Table 5.3).

During cell division, homologous chromosomes (meiosis I) or sister chromatids (meiosis II, mitosis) segregate to opposite poles, ensuring proper distribution of genetic material to each daughter cell. The centromere is the site of formation for the kinetochore, which interacts with microtubules of the spindle apparatus to permit chromosome alignment during metaphase and segregation during anaphase. It is thought that centromeric heterochromatin (consisting of centromeric DNAs and proteins) facilitates the nucleation of the kinetochore. The work presented in this thesis furthers our knowledge on the constitution, genesis, and evolution of mammalian centromeric DNAs and will hopefully enhance our understanding of the requirements and mechanisms of proper chromosome segregation.

References:

Bogenberger JM, Neumaier PS, Fittler F. (1985) The muntjak satellite IA sequence is composed of 31-base-pair internal repeats that are highly homologous to the 31-base-pair subrepeats of the bovine satellite 1.715. *Eur J Biochem* 148: 55-59.

Bogenberger JM, Neitzel H, Fittler F. (1987) A highly repetitive DNA component common to all cervidae: its organization and chromosomal distribution during evolution. *Chromosoma* 95: 154-161.

Choo KH, Vissel B, Brown R, Filby RG, Earle E. (1988) Homologous alpha satellite sequences on human acrocentric chromosomes with selectivity for chromosome 13, 14, and 21: implications for recombination between non homologues and Robertsonian translocations. *Nucleic Acids Res* 16: 1273-1284.

Haaf T, Steinlein K, Schmid M. (1986) Preferential somatic pairing between homologous heterochromatic regions of human chromosomes. *Am J Hum Genet* 38: 319-329.

Lee C, Ritchie DBC, Lin CC. (1994) A tandemly repetitive, centromeric DNA sequence from the Canadian woodland caribou (*Rangifer tarandus caribou*): its conservation and evolution in several deer species. *Chromosome Res* 2: 293-306.

Lin CC, Sasi R, Fan Y-S, Chen Z-Q. (1991) New evidence for tandem chromosome fusions in the karyotypic evolution of Asian muntjacs. *Chromosoma* 101: 19-24.

Lin CC, Sasi R, Lee C, Fan YS, Court D. (1993) Isolation and identification of a novel tandemly repeated DNA sequence in the centromeric region of human chromosome 8. *Chromosoma* 102: 333-339.

Masumoto H, Masukata H, Muro Y, Nozaki N, Okazaki T. (1989) A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *J Cell Biol* 109: 1963-1973.

Scherthan H. (1991) Characterization of a tandem repetitive sequence cloned from the deer *Capreolus capreolus* and its chromosomal localisation in two muntjac species. *Hereditas* 115: 43-49.

Willard HF, Waye JS. (1987) Hierarchical order in chromosome-specific human alpha satellite DNA. *Trends in Genet* 3: 192-198.

Yu LC, Lowensteiner D, Wong EFK, Sawada I, Mazrimas J, Schmid C. (1986) Localization and characterization of recombinant DNA clones derived from the highly repetitive DNA sequences in the Indian muntjac cells: Their presence in the Chinese muntjac. *Chromosoma* 93: 521-528.